THE DEVELOPMENT, DESIGN AND TESTING
OF A PERFORMANCE MEASUREMENT MODEL
FOR A NOT-FOR-PROFIT ORGANIZATION

BY

W. MARCUS DUNN

A DISSERTATION PRESENTED TO THE GRADUATE
COUNCIL OF THE UNIVERSITY OF FLORIDA IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA
1976

To

Diane and Kristen

for Their Love and Patience

During the Interment of

"Oscar the Grouch"

TABLE OF CONTENTS

TABLE OF CONTENTS—Continued

TABLE OF CONTENTS--<u>Continued</u>

LIST OF TABLES

LIST OF TABLES--Continued

LIST OF FIGURES

THE DEVELOPMENT, DESIGN AND TESTING
OF A PERFORMANCE MEASUREMENT MODEL
FOR A NOT-FOR-PROFIT ORGANIZATION

By

W. Marcus Dunn

August, 1976

Chairman: D. D. Ray
Major Department: Accounting

Introduction and Purpose

Although expenditures by not-for-profit (NFP) organizations
account for over one-third of the gross national product of the
United States, the accounting systems for such organizations provide
little more than budgetary control over expenditures. Furthermore,
despite recognition of the need for performance measurement, oper-
ational performance measurement systems are virtually nonexistent.

The purpose of this research was to develop and test a perform-
ance measurement model for a NFP organization. The research
proceeded in three phases. First, the present state of performance
measurement in the NFP area was assessed. Second, the model was
developed. Finally, the model was operationalized by collecting
and analyzing the data required by the model.

Model Development and Test

The model was developed for the Gainesville Recreation Department (GRD), Gainesville, Florida. For the purpose of model development, the GRD was viewed as a social system whose purpose is to promote the welfare of the Gainesville community through the provision of recreation programs and facilities. Ideally, the amount and type of public recreation provided would be the amount and type consistent with the maximum total welfare for the Gainesville community. Through knowledge of the social welfare function for Gainesville, decision makers establish the optimal recreation budget, acquire program inputs and transform the inputs via production functions into those outputs consistent with maximum welfare.

Because the social welfare function exceeds current knowledge, operationalization of the model was limited to the development and validation of measures of program output and input. With the assistance of GRD supervisors, major recreation programs were identified. For most of these programs, the GRD was able to provide estimates of the following objective inputs and outputs: (1) direct costs, (2) labor hours of input, (3) participant and spectator hours (quantity of output) and (4) user fees. As measures of output quantity, by themselves, provide insufficient evidence of how well the community is served by a program, measures of program importance and quality were also obtained. The measures of program importance and quality were produced from the opinions of GRD supervisors, Public Recreation Advisory Board members and a sample of residents in the Gainesville

xii

community. The validity of these subjective measures was evaluated
by use of Delphi procedures and multitrait-multimethod methodology.

Conclusions

Based on Delphi criteria, valid group judgements of program
importance and quality were found to exist. Based on multitrait-
multimethod criteria, the measures of program importance and quality
are believed to be valid. The success achieved in measuring and val-
idating the importance and quality of recreation programs suggests
that subjective measures are well within the purview of current
methodology--to the extent such measures are found useful to NFP decision
makers, they can and should be provided. While a lack of research
resources precluded the measurement of objective inputs and outputs
as well as desired, the generation process itself has demonstrated the
potential for formally collecting such information.

The successful collection and validation of the data specified
by the model indicates that the model can be operationalized. Its
implementation would provide quantitative information to decision
makers which should be useful in assessing the contribution of individ-
ual recreation programs to the welfare of the Gainesville community
and in evaluating the impact of budget changes on outputs and inputs.

A. H. Ray
Chairman

CHAPTER I

INTRODUCTION

Purpose of Research

The general purpose of this research is to provide evidence as
to the feasibility and efficacy of performance measurement in the
not-for-profit (NFP) area. The researcher hopes to accomplish this
through the design, development and testing of a performance measure-
ment model for a NFP organization. Procedurally this will involve (1)
a review and assessment of existing methodologies for NFP performance
measurement, (2) the design and development of a model for measuring
the performance of the Gainesville Recreation Department (GRD) and
(3) the operationalization of this model.

Contribution of Research

Because little theory or prior research esists to guide this
research project, it must be viewed as exploratory. This fact does
not render it any less valuable, however, for as Bauer has noted

> ...it is recognized that at the present state of the art,
> the first efforts to collect new kinds of data will be
> seriously defective. Here the conclusion seems to be that
> rather than do nothing it is preferable to start out with
> bad data, warn everyone about the defects and limitations
> and aim at gradual "improvement through use." [emphasis
> added] (1967, p. .vi)

Thus while definitive conclusions should not be expected, the following
contributions to knowledge are anticipated:

1. Although concern for performance measurement in the NFP area

1

is evident, researchers are handicapped by the fact that the present
state of this type of measurement is not well known. In this re-
search, methodologies which have been proposed for dealing with
performance measurement in the NFP area will be reviewed and eval-
uated. The extent of their applications and accomplishments will
be discussed. This information should be of value to others who
desire to do research in the area.

    2. Although the need for performance measurement is well
recognized, few attempts have been made to develop formal models
for assessing the performance of NFP organizations. In this research
a performance measurement model will be developed for the GRD. The
model development and test should provide valuable insight into the
problems and benefits of a formal evaluation system. The model will
hopefully serve as a guide and catalyst for the future development
of performance measurement models.

    3. Evaluation of the performance of recreation departments
has been limited primarily to self-evaluation studies (e.g., Smissen,
1972). Because the objectivity of such evaluations is questionable,
this research utilizes both self and outside evaluations. By comparing
these evaluations, conclusions as to the validity (and therefore the
usefulness) of the self-evaluation study should be possible.

    4. Helmer (1966) and Dalkey (1969) advocate the use of expert
judgement via the Delphi technique as a means of dealing with areas
in which exact knowledge is not available. Dalkey has stated that
there are two options available when one is working on a problem

under conditions of uncertainty with insufficient data, incomplete theory, and a high order of complexity:

> ...we can either wait indefinitely until we have an adequate theory enabling us to deal with socio-metric and political problems as confidently as we do with problems in physics and chemistry, or we can make the most of an admittedly unsatisfactory situation and try to obtain the relevant intuitive insights of experts and then use their judgements as systematically [via Delphi] as possible. (quoted in Pill, 1971, p. 61)

In the development of a social service measurement model for the Cleveland Jewish Community Federation (Mantel et al., 1972), Delphi was found to be a useful means of producing measures of service quality and value. Its use in this research to generate measures of the importance and quality of recreation programs should provide additional evidence as to its usefulness in developing performance measures.

5. The multitrait-multimethod methodology (Campbell and Fiske, 1959) is a powerful tool for assessing the validity of constructs incapable of exact measurement. While this methodology has been applied in the area of managerial performance measurement (Lawler, 1967), its use in this research represents its first known application for the purpose of evaluating measures to be used in assessing the performance of a NFP organization. This methodology will hopefully provide a means of objectively appraising subjective measures.

Scope of Research

The literature reviewed and evaluated in this research deals primarily with NFP organizations in the United States. The design, development and testing of the performance measurement model was done

with the cooperation and assistance of the Recreation Department of
the City of Gainesville, Florida.

Nature of the Not-For-Profit Area

The fundamental distinction between profit and NFP organizations
is that the primary goal of profit organizations is to maximize profits
whereas a NFP organization is concerned only indirectly, if at all,
with profit maximization. NFP organizations, such as governments,
schools, churches and hospitals, are primarily concerned with the
welfare of their beneficiary groups. These organizations do not sell
their products and services in the marketplace and therefore the
benefits they provide can seldom be measured in terms of revenue.

Historically, the primary indicant of the performance of profit
organizations has been the net income generated for some past period
of time. Net income, a product of a firm's accounting system, is the
result of applying the matching concept to economic events affecting
the firm. Since revenues are not normally produced or sought by NFP
organizations, net income cannot be used to evaluate performance in
the NFP area (e.g., Knighton, 1969; Committee on Concepts of Accounting
Applicable to the Public Sector, 1972).

Importance of Not-For-Profit Area

Presently expenditures by NFP organizations constitute over one-
third of the gross national product of the United States. Expenditures
by government, the largest type of NFP organization, have increased
from 7% of GNP in 1902 to 34% of GNP in 1970 (Lee and Johnson, 1973,
p. 38). On a per capita basis this growth represents an increase

from $90 (in 1970 dollars) in 1900 to $1,638 in 1970 (Ibid., p. 39).
Of course, government is not the only NFP organization which has
experienced significant spending increases. For example, expendi-
tures by hospitals increased from $3.7 billion in 1950 to $28.8
billion in 1971 (American Almanac, 1974, p. 75) and expenditures by
schools increased from 8.8 billion in 1950 to $90.2 billion in 1973
(Ibid., p. 108). On a per capita, real basis (1950 expenditures
expressed in 1971 and 1973 dollars respectively), the respective
increases were from $33.82 to $139.10 (hospitals) and from $92.33
to $428.69 (schools).

In view of the magnitude of expenditures in the NFP area and
the scarcity of resources, it is imperative that available resources
be used efficiently and effectively. The measurement of performance
has never appeared to be more critical.

Lack of Performance Measures in the Not-For-Profit Area

The lack of performance measures in the NFP area is well docu-
mented in the literature. Bauer states that "for many of the impor-
tant topics on which social critics blithely pass judgement, and on
which policies are made, there are no yardsticks by which to know
if things are getting better or worse" (1967, p. 20). Terleckyj notes
that "in contrast to a few exceptionally advanced fields most areas
of social concern and public policy suffer from lack of even the most
elementary information, leaving the field wide open for guesswork,
emotion, low grade politics and waste" (1970, p. B-765). Wholey
et al. write that

> The most impressive finding about the evaluation of
> social programs in the federal government is that sub-
> stantial work in this field has been almost nonexistent....
> There is nothing akin to a comprehensive federal eval-
> uation system. Even within agencies, orderly and inte-
> grated evaluation operations have not been established.
> (1970, p. 15)

The Committee on Not-For-Profit Organizations asserted that "no 'general government' performance criteria, indicators and/or control devices analogous to the net income measurement of business type activities have been developed to date" (1974, p. 228).

That the present state of affairs is disconcerting to the highly educated citizenry characteristic of modern industrial societies is evident from the following comments by Churchman:

> There is no question that in our age there is a great
> deal of turmoil about the manner in which our society is
> run. Probably at no prior point in the history of man
> has there been so much discussion about the rights and
> wrongs of the policy makers....In all cases the citizen
> feels a perfect right to have his say about the way in
> which the managers manage....
>
> Not only has the citizen become far more vocal, but he
> has also in many instances begun to suspect that the
> people who make the major decisions that affect our
> lives don't know what they are doing. They don't know
> what they are doing simply because they have no adequate
> basis to judge the effects of their decisions. (1968, p. vii)

Extending the Role of Accounting

The accountant has thus far limited his involvement in the NFP area to providing control over inputs--accounting systems at present only indicate compliance with legal and budgetary restrictions. These systems provide little information as to the accomplishments of organizational objectives or the efficient use of resources (e.g., Henke, 1972; Committee on Accounting Practices of Not-For-Profit

Organizations, 1971; Committee on Concepts of Accounting Applicable
to the Public Sector, 1972). With the recognition of the need for
performance measures in the NFP area, some accountants have advo-
cated an extension of the accountant's role beyond mere dollar
accountability. Among those advocating an increased role by the
accounting profession in evaluating performance are Bedford (1962),
Churchill and Stedry (1967), Beyer (1969), Knighton (1969 and 1972),
Mobley (1970), Estes (1972), Henke (1972), Brummett (1973) and
Linowes (1973). As to the significance of such an extended role,
Bedford has written that "whoever discovers a means for measuring
objectively the accomplishment of such organizations [NFP] will
contribute greatly to the accounting discipline" (1962, p. 93).

Knighton, one of the strongest advocates of extending the
role of accounting, believes that the following action is required:

> We are not in business to spend the public funds.
> Rather, we are charged with the responsibility for
> using public resources to bring about conditions that
> promote the public good. And as accountants in this
> all-important endeavor, we must become as concerned with
> accomplishment of this mission as with accounting for the
> resources used. (1972, pp. 5-6)

Furthermore, he states

> The objectives of government is not to earn revenue. The
> objective of government is to provide benefits to society.
> And if effort and accomplishment are to be compared or
> related here, we must match operating expenses with infor-
> mation that gives us an indication of public benefits.
> Only then can we really say something conclusive about
> efficiency and effectiveness in government operations....
>
> We must learn to design and operate systems that report
> information on outputs and accomplishments. [Emphasis
> added] (1972, pp. 7-8)

However, recommendations for the extension of the accountant's scope beyond financial accounting are not limited to members of the accounting discipline. Among others, Lazarsfeld (1971), a sociologist, and Churchman (1971), a management science philosopher, have urged the accountant to serve society better by moving beyond financial accounting into social accounting.

Although the exact role the accountant will play in the development of performance measures is presently unsettled, there are indications of a willingness on the part of certain members of the accounting profession to move beyond mere dollar accountability. For example, a survey of big eight C.P.A. firms by the Committee on Measures of Effectiveness for Social Programs revealed considerable enthusiasm for undertaking social program evaluation engagements (1972, p. 387). Another example is provided by the General Accounting Office which has become active in evaluating the efficiency and effectiveness of federal agencies. Its new Standards for Audit of Governmental Organizations, Programs, Activities and Functions represents a significant movement beyond financial accounting:

> This demand (from public officials, legislators and the
> general public) for information has widened the scope of
> governmental auditing so that such auditing no longer is
> a function concerned primarily with financial operations.
> Instead, governmental auditing now is also concerned with
> whether governmental organizations are achieving the
> purpose for which programs are authorized and funds made
> available, are doing so economically and efficiently and
> are complying with applicable laws and regulations. (1972,
> p. i)

Although the need is recognized, thus far performance measurement in the NFP area has been little more than an attractive slogan.

Operational performance measurement systems are virtually non-existent. The accounting literature specifically and the social science literature in general reveal activity primarily at a conceptual, normative level. Empirical studies are needed in order to develop and refine performance methodologies and to provide evidence on the feasibility and efficacy of performance measurement in the NFP area. The failure of such research to be forthcoming will probably result in a dissipation of the energy and enthusiasm which have been generated by the numerous heuristic expressions of interest in performance measurement. The warning of the Committee on Accounting for Human Resources appears especially applicable to the present state of development of performance measurement systems:

> Perhaps the most important task facing those who wish to do advance work in accounting for human resources [substitute performance measurement] stems from the necessity to demonstrate the usefulness of HRA [performance measurement] systems. Unless empirical data from organizations using HRA systems are collected, analyzed and published, the attractiveness of current theoretical arguments for HRA may soon lose their glamour. (1974, p. 124)

Chapter Preview

In this chapter the purpose of and need for this research was set forth.

In Chapter II a synthesis of the present state of performance measurement in the NFP area will be presented. This synthesis entails (1) an examination of methodologies which have been proposed for use in measuring performance in the NFP area and (2) a discussion of the extent to which these methodologies have been applied. The

usefulness of these methodologies for this research will also be discussed.

In Chapter III the performance measurement model developed for evaluating the performance of the GRD is presented and the specific procedures required to operationalize this model are delineated.

In Chapter IV the methodologies to be employed in developing and testing the model are discussed.

In Chapter V the results of collecting and analyzing the data specified by the performance measurement model will be presented.

In Chapter VI the conclusions reached will be stated. Problems encountered and recognized deficiencies will be discussed. The applicability of the model to entities other than recreation departments will be assessed. Recommendations for future research will be proposed.

CHAPTER II

THE PRESENT STATE OF PERFORMANCE MEASUREMENT
IN THE NOT-FOR-PROFIT AREA

Introduction

As indicated in Chapter I performance measurement in the NFP
area is in its infancy. Although the accountant has been directly
involved with NFP organizations for many years, he has almost totally
neglected the area of performance measurement.

Fortunately, members of other disciplines have attempted to
develop methods for measuring the performance of NFP activities.
It is therefore appropriate that we examine these methods in order
(1) to indicate what has previously been done in the area and there-
by present a clearer view of the current state of performance measure-
ment and (2) to assess the usefulness of the various methods for
the design and development of a performance measurement model for
the GRD. The methods to be examined are

1. Pareto-optimality

2. Cost-benefit analysis

3. Cost-effectiveness analysis

4. Planning-Programming-Budgeting

5. Social indicators

6. Experimental and quasi-experimental research designs

Because of its contribution to this research effort, the devel-
opment of a social service measurement model for the Cleveland

11

Jewish Community Federation (Mantel et al., 1972) will also be dis-
cussed. Finally, political rationality, a school of thought crit-
ical of the formal, rational methods, will be examined.

Pareto-Optimality

One discipline which has long been concerned with optimal
performance is economics. Whether from the point of view of the
individual, the firm or society, economists have sought to determine
how resources can be allocated so as to produce maximum welfare.

The branch of economics which takes the broad social pers-
pective is welfare economics, whose objective is "the evaluation
of the social desirability [emphasis added] of alternative economic
states" (Hendersen and Quandt, 1971, p. 254), The desired state
is the one for which social welfare is at a maximum.

This optimal state cannot be identified, however, because the
utilities of individuals are not comparable and therefore there is
presently no way to meaningfully relate them in a single social
welfare function. As a result welfare economics has come to be
based primarily on the concept of Pareto-optimality (Hendersen and
Quandt, 1971, p. 255).

A state is Pareto-optimal if no reorganization would result
in someone being better off without making someone else worse off.
If a state is not Pareto-optimal, it is possible to reorganize it
with a concomitant increase in social welfare. Thus Pareto-optimal-
ity can serve as a social welfare criterion.

Unfortunately, its usefulness as such a criterion has proved very limited because

1. it is extremely difficult, if not impossible, to assess the impact of a reorganization on the utility of each member of society

2. most reorganizations result in some of society's members being made better off while others are made worse off

3. utility, itself, is an abstract concept which has never been successfully quantified.

To make Pareto-optimality more useful, the compensation principle was introduced (Henderson and Quandt, 1971, p. 279). According to this principle, a change, in which gainers would be able to compensate the losers so that all individuals in society would either favor the change or be indifferent to it, is desirable. However, since compensation principles generally refer to potential rather than actual compensation, "the compensation criteria imply interpersonal comparisons of utility that most economists strive to avoid" (Ibid., p. 280).

Welfare economics in general and Pareto-optimality in particular are presently of limited value to those desiring to assess the performance of NFP entities. According to Cohn, "the area of welfare economics cannot offer highly optimistic grounds for the choice of ideal policy criteria that are designed to maximize welfare" (1972, p. 45). As a result, decision makers have had to turn to sub-optimal techniques, such as cost-benefit analysis, for guidance.

Cost-Benefit Analysis

Definition and Purpose

Cost-benefit analysis, as an aid to optimal performance in the NFP area, is a method of evaluating and ranking in terms of economic desirability projects posited to be fruitful in attaining certain social goals. Implementation of a cost-benefit analysis requires for each project

1. a specification of all costs and benefits

2. a valuation of the specified costs and benefits

3. a determination of an appropriate discount rate

4. specification of relevant constraints (Prest and Turvey, 1965, p. 158).

Using the above information the net present value of a project can be computed. A project merits consideration only if its net present value is positive. (A negative net present value implies that the social opportunity cost exceeds the social value of the benefits.) If several projects are being considered, they can be ranked in terms of their net present value which provides the criterion for choosing the most desirable project--the one with the largest net present value. Ideally, if all projects being considered were perfectly divisible, resources could be allocated among the various projects so that the ratio of marginal benefits to marginal costs were equal for all projects.

In addition to a positive net present value, if cost-benefit analysis is to serve as an _ideal_ welfare criterion, certain stringent conditions must be satisfied (Krutilla, 1966). These are

1. opportunity costs are borne by beneficiaries in such wise as to retain the initial income distribution

2. the initial income distribution is in some sense 'best'

3. the marginal social rates of transformation between any two commodities are everywhere equal to their corresponding rates of substitution except for area (s) justifying the intervention in question. (Krutilla, 1966, p. 177)

In addition to limitations produced by the absence of the preceding conditions, significant practical implementation problems also exist:

1. It is often difficult to identify all relevant benefits and costs and even where they can be identified, quantification in dollar terms may not be possible.

2. Future benefits and costs are uncertain.

3. Net present values are often quite sensitive to the discount rate chosen--presently agreement does not exist as to how an appropriate discount rate is to be selected.

4. The need to adjust market prices for market imperfections and anticipated price changes introduces considerable uncertainty into the analysis.

5. Market prices cannot be used to value benefits which cannot be marketed--the collective goods problem.

In view of such difficulties, the results of a cost-benefit analysis will generally be imprecise. According to Dorfman, "The debate about benefit-cost analysis centers on the question of whether the social value of benefits can be estimated reliably enough to justify the trouble and effort involved in a benefit-cost computation" (1965, p. 8).

Despite its (cost-benefit analysis) limitations, Krutilla believes
that

> Since the alternative is not to retire to inactivity but,
> rather, to reach decisions in the absence of analysis, we
> may take some comfort from the belief that thinking syste-
> matically about problems and basing decisions on such analysis
> are likely to produce consequences superior to those that
> would result from purely random behavior. (1966, p. 189)

Prest and Turvey, while fully cognizant of the problems of cost-benefit

analysis, also see certain advantages in its application:

> 1. ...it forces those responsible to quantify costs and
> benefits as far as possible rather than rest content with
> vague qualitative judgements or personal hunches....
>
> 2. ...it has the very valuable by-product of causing ques-
> tions to be asked...which would otherwise not have been
> raised,....
>
> 3. ...even if cost-benefit analysis cannot give the right
> answers, it can sometimes play the purely negative role in
> screening projects and rejecting those answers which are
> obviously less promising. (1965, p. 202)

Applications

That cost-benefit analysis is not a new idea is evident from

an examination of the River and Harbor Act of 1902 which required

that the desirability of river and harbor projects be determined by

consideration of the amount of commerce benefited and the cost (Prest

and Turvey, 1965, p. 155). However, it is only in recent years that

the technique has received widespread interest and attention. Prest

and Turvey attribute the heightened interest to (1) the growth of

large investment projects, (2) the growth of the public sector and

(3) the development of operations research and systems analysis (1965,

p. 156).

Recalling that for the successful implementation of cost-benefit analysis, benefits and cost must be both identified and valued, it is understandable that cost-benefit analysis has been most extensively applied to the areas of water resources and transportation. Projects in these areas (irrigation, hydroelectric, flood control, road, railway and inland waterway) generally yield some directly identifiable outputs (benefits) which are usually valued in the marketplace. Cost-benefit analysis has also been applied, on a much more limited scale, to the areas of land usage, health, education, research and development and defense. For areas such as these, however, it is much more difficult to identify the benefits produced by a particular project. Furthermore, the benefits are often not valued in the private marketplace. Cost-benefit ratios produced for projects in these areas will generally be more unreliable than those for the areas of water resources and transportation.

Use In This Research

Although cost-benefit analysis appears to be a useful method for evaluating projects in certain areas, it is of dubious value for this research. For recreation in general, the relationships between recreational activities and social well-being, physical and mental health, productivity, crime, property values and economic growth have not been empirically established. More specifically, data on the assumed benefits provided the Gainesville community by the City's recreation programs do not exist and market values for

these benefits are virtually non-existent. In view of these facts,
reliable cost-benefit ratios could not be produced.

Furthermore, unless opportunity costs are borne by beneficiaries,
the provision of public recreation programs will alter the distri-
bution of income in the Gainesville community. Now since most of
the expense of recreation is paid from ad valorem taxes and not from
recreation fees based on usage of programs and facilities, it is
very unlikely that opportunity costs are being borne by beneficiaries.
For example, upper income groups, who have more wealth but less
need for public recreation, would appear to be subsidizing lower
income groups, who have greater need but less wealth. Such distri-
bution effects limit the usefulness of cost-benefit analysis as an
evaluation tool.

Cost-Effectiveness Analysis

Definition and Purpose

Since cost-benefit analysis requires market prices for valu-
ation of benefits, it is not applicable in areas where such prices
do not exist. To provide a guide to rational decision making in
these areas, a technique entitled cost-effectiveness analysis has
been developed. Utilizing this technique, it is possible to maxi-
mize a certain output (non-dollar) per a constant cost or to mini-
mize the cost of producing a constant output.

In order to implement a cost-effectiveness analysis, project
outputs must be specifiable and project costs must be capable of

valuation at market prices.

Since future outputs and costs are uncertain, the results of
a cost-effectiveness analysis will be imprecise and must be inter-
preted with caution. Because outputs are not valued, the technique
cannot provide an index of social desirability (Cohn, 1972, p. 73).
Furthermore, the absence of the dollar as a common metric, means
that the technique cannot be used to guide choices between different
goals.

Applications

Althought the ultimate origins of cost-effectiveness analysis
lie in economic production and utility theory, its birth as an analyt-
ical technique must be assigned to the efforts of defense research
during and after World War II (Goldman, 1967, p. v). As might be
expected, its greatest use has been in the area of national defense.

Use In This Research

Since cost-effectiveness analysis does not require the valu-
ation of benefits, it has potential for use in this research project,
provided the output of the GRD can be defined in some meaningful
way.

Planning-Programming-Budgeting

Definition and Purpose

Planning-Programming-Budgeting (PPB) is an approach to budgeting
which seeks to extend the budgetary process beyond the area of input
control and into the area of efficient and effective resource manage-

ment. It focuses on the program as the basis for planning and bud-geting—the program being the total resource commitment for achieving some common objective. The application of PPB entails (1) setting goals, (2) determining the alternative programs for achieving these goals, (3) using some decision rule (such as provided by a cost-benefit or a cost effectiveness analysis) to select the optimal program, and (4) allocating resources to the optimal programs.

PPB represents a significant departure from the traditional line item method of budgeting. Whereas in traditional budgeting the guide to current resource allocation is the previous year's budget, the guide to resource allocation in a PPB approach is the goal to be accomplished and the program resources thought to be required for achieving this goal.

Another significant and important difference between PPB and traditional budgeting is that under PPB the time period considered extends beyond the single budget year. PPB involves multi-year planning and a multi-year budget.

Applications

During the 1950's, the Rand Corporation developed PPB for use in defense budgeting. In 1961 under Defense Secretary Robert McNamara's direction and support PPB was introduced into the Department of Defense. McNamara required that the defense budget for fiscal 1963 be formulated in terms of major programs and weapon systems (Held, 1970, p. 13). "The results of this reorganization and of the evalu-ations it has made possible, led to recommendations that the approach

be extended to civilian affairs" (Ibid., p. 13).

In 1965 President Lyndon Johnson announced that all federal
agencies would be required to use PPB. According to the President,
each federal department was required to

Develop its objectives and goals, precisely and carefully;

Evaluate each of its programs to meet these objectives,
weighing the benefits against the costs;

Examine in every case, alternative means of achieving
these objectives;

Shape its budget request on the basis of this analysis
and justify that request in the context of a long range
program and financial plan. (Lyden and Miller, 1970, p. 5)

Given all the fanfare and optimism with which PPB was introduced into
the federal government*, it surely must have appeared to some that
a new era of rationality had engulfed the public expenditure process.
This euphoria, however, was to be short lived. The difficulties
of translating a conceptual PPB into a functioning viable PPB were
to prove insurmountable.

Schick, in a review of PPB, noted that

The publicity has outdistanced the performance by a wide margin.
In the name of analysis, bureaus have produced reams of un-
supported, irrelevant justification and description. As
Shumpeter said of Marxism: it is preaching in the garb of
analysis. (1969, p. 149)

Stanley Botner in another review of PPB reported that

---

*Although a few state and local governments have introduced PPB into
their budget systems, PPB should be viewed primarily as a federal
effort.

> In mid-1968 the Bureau of the Budget undertook to determine
> if policy analysis is performed differently than it was be-
> fore the advent of PPB. BOB researchers conducted a study
> of implementation and utilization of PPB by 16 domestic
> federal agencies. Their findings led them to conclude that
> most agencies do not perform the planning, programming, and
> budgeting functions much differently than they did before
> the introduction of PPB. (1970, p. 423)

Furthermore, "An analysis of the results of recent studies, dis-

cussions with Budget Bureau and other officials and testimony before

congressional subcommittees leads one to conclude that PPB has thus

far been rather ineffectual as a presidential staff tool" (Ibid.).

In view of its lack-luster performance, it was not surprising

that a change in administration's (from Johnson to Nixon) sealed

the doom of PPB. In June, 1971 a memorandum from the Office of

Management and Budget stated that

> Agencies are no longer required to submit with their budget
> submissions the multi-year program and financing plans, pro-
> grams memoranda and special analytical studies.., or the
> schedules... that reconcile information classified according
> to their program and appropriation structures. (Office of
> Management and Budget Transmittal Memorandum No. 38, June
> 21, 1971)

"By these words," according to Schick "PPB became an unthing"

(1973, p. 146).

Post mortems as to the cause of death (e.g., Schick, 1973;

Tiller, 1972) offer the following reasons for PPB's failure:

1. PPB was introduced on too large a scale and without suffi-

cient advance preparation and planning.

2. PPB was suddenly thrust upon the various federal agencies

without their participation in the decision to implement PPB. Many

agency leaders were thus initially alienated from PPB and never became committed to it.

3. There never existed a sufficient number of personnel capable of providing the information and analysis needed for a total federal PPB. At its peak the Bureau of the Budget staff responsible for PPB in all the federal agencies numbered less than 12.

4. The data needed for various analytical purposes simply did not exist.

5. PPB ignored the traditional budgeting process—yet Congress and most of the federal agencies continued to use this process as the one for decision making.

It does not appear that PPB failed because of conceptual unsoundness. Rather its failure must be attributed to deficiencies in implementation, disregard of the traditional budgeting process and a lack of sufficient resources for administration. Although PPB, as a total federal effort, has ceased to exist, it has served to stimulate a great interest in program evaluation and economic rationality.

Use In This Research

The City of Gainesville uses traditional, line item budgeting. A system comparable to PPB has never been attempted and it is well beyond the scope of this research project to attempt to install a PPB system for the GRD.

However, it appears that the program is a practical and useful unit of account. Therefore major recreation programs will be identified and measures of program input and output will be produced.

Social Indicators

Definition and Purpose

According to Bauer, social indicators are "statistics, statis-
tical series, and all other forms of evidence that enable us to assess
where we stand and are going with respect to our values and goals
and to evaluate specific programs and determine their impact" (1967,
p. 1). Thus defined, economic indicators are a subset of social
indicators. However, it was dissatisfaction with the narrowness
of economic indicators (e.g., gross national product, unemployment)
that has fostered so much interest in the broader concept of social
indicators;

> ...the highly quantitative economic data in today's econ-
> omic survey documents tend to detract attention from ideas
> that cannot be so readily expressed in quantitative terms....
> Economic statistics, as a whole, emphasize the monetary value
> of goods and services. By so doing, they tend to discrimi-
> nate against nonmonetary values and against public services
> for which costs invariably serve as surrogates of output
> value. Because figures on health and life expectancy are
> not directly incorporated in national economic accounts,
> progress in these areas may be seriously ignored, either
> in formulating goals or in evaluating performance....
>
> In short, national economic accounting has promoted a
> "new Philistinism" -an approach to life based on the prin-
> ciple of using monetary units as the common denominator of
> all that is important in human life.
> ....
> This bias can be overcome only by persistent efforts to
> develop broader models that include many more variables than
> those thus far used by economists. (Gross, 1967, pp. 167-168)

The consensus of opinion of the advocates of systems of social indi-
cators is that economic indicators, while necessary, are not suf-
ficient for the evaluation of social systems and the measuring of

progress towards social system goals.

Some of the areas which have been identified as being extremely important to society's welfare but which are not measured by economic indicators are (1) health, (2) social mobility, (3) condition of physical environment and human habitat, (4) public order and safety, (5) learning, science, art, leisure, and (6) freedom and justice. Concerning indicators of progress in such areas as those noted above, the following statements from Toward a Social Report are apropos:

> The Nation [United States] has no comprehensive set of statistics reflecting social progress or retrogression. There is no Government procedure for periodic stocktaking of the social health of the Nation. The Government makes no Social Report [emphasis added].(1969, p. xi)

Applications

Although comprehensive systems of social indicators have been formulated (e.g., Gross, 1967; Terleckyj, 1969), the transition from a conceptual to an operational system remains. At the conceptual level, the problems of (1) goal specification and agreement, (2) identification and validation of measures of goal attainment, and (3) discovering the relationships between inputs (resources used) and outputs (posited as conducive to goal attainment) can be ignored or assumed away. But at the operational level, these problems must be addressed successfully. This will require great effort, expense and time.

Use In This Research

In this research project an attempt will be made to develop

and validate indicators believed useful for assessing the performance
of the GRD. In a sense, the model to be developed is an information
system which provides indicators of program cost, usage, quality
and importance.

Experimental and Quasi-Experimental
Designs for Program Evaluation

Definition and Purpose

The current interest in program evaluation has resulted in
an advocacy by some social scientists of the use of experimental
and quasi-experimental research designs* for the purpose of deter-
mining program effectiveness (e.g., Campbell, 1969; Rossi and
Williams, 1972; Campbell, 1974). According to Campbell

> The United States and other modern nations should be ready
> for an experimental approach [emphasis added] to social
> reform, an approach in which we try out new programs designed
> to cure specific social problems, in which we learn whether
> or not these programs are effective, and in which we retain,
> imitate, modify or discard them on the basis of apparent
> effectiveness on the multiple imperfect criteria available.
> (1969, p. 409)

True experimental design form the basis for much of the "experimental
approach" referred to by Campbell. Such designs require

1. manipulation of independent variable(s) by an investi-
gator (such manipulation may take the form of exposing one group
(the treatment group) to a particular social program while with-

---

*For a comprehensive discussion of experimental and quasi-experimental
research designs, along with the various threats to internal and ex-
ternal validity, the reader may consult Campbell and Stanley, 1963.

holding it from another (the control group) or it may mean exposing each group to a different program)

    2. establishing statistical equivalency between treatment and control groups by randomly assigning a pool of subjects (ideally selected at random from a larger population) to each group

    3. observing and measuring the variance between the treatment and control groups.

The great power of the experimental design lies in its control of alternative explanations (rival hypotheses) of the cause of the variance observed between the treatment and control groups. Unless these alternative explanations are eliminated, the variance observed cannot be warrantedly attributed to the treatment (social program) and the effectiveness of the treatment will remain unknown.

    Situations often exist for which equivalent treatment and control groups are not possible. Therefore one of the basic requirements for the experimental design is lacking. In such situations, however, it is often possible to achieve something comparable to the experimental design by utilizing non-random control groups or by using the treatment group as its own control. Such research designs are referred to as quasi-experimental designs. Since for these designs, full experimental control is lacking, it is essential that the researcher be fully aware of the alternative explanations as to the cause of observed variance.

Applications

    Although the canon of controlled, comparative experimentation

was established by Fisher in the 1920's, subsequent experiments
in the Fisherian tradition have been limited primarily to the labor-
atory and the agricultural experiment station (Campbell, 1969, p. 425).
Rossi, while noting that "there exist elegant models for carrying
out evaluation studies, derived mainly from the controlled experiment
tradition," finds that "there are almost no examples of evaluation
studies of current programs which have followed these models with
any appreciable degree of fidelity" (1972, p. 29).

The greatest obstacle to the use of experimental designs for
evaluating social programs appears to be the reluctance of program
administrators to allow subject to be assigned, at random, to the
treatment and control groups. Randomization is seen as inhumane
and it conjures up notions of eccentric scientists.

Another difficulty lies in the fact that for some programs
randomization is not possible either because potential subjects are
limited in number or because the program is an all or nothing affair
(not possible to have a control group).

Such problems in utilizing the true experiment have resulted
in recommendations that the quasi-experimental designs be used where
it is impractical or impossible to utilize the experimental designs
(e.g., Campbell and Stanley, 1963; Campbell, 1969).

Use In This Research

Experimental and quasi-experimental designs are powerful methods
for determining the impact and effectiveness of programs. However,
their use requires direct intervention with presently existing pro-

cesses. Since intervention with the current recreation programs

of the GRD is outside the scope of this research project, experi-

mental and quasi-experimental designs are inappropriate for this research.

A Social Service Measurement Model for the
Cleveland Jewish Community Federation

Interest in measurement models for the evaluation of NFP organ-

izations has been exhibited in recent years. One model of particular

interest and merit was developed by an interdisciplinary research

team at Case Western Reserve University. The research team undertook

(1) "to develop a consistent, relevant and reasonably reliable set

of data on the services offered" by the 21 agencies of the Cleveland

Jewish Community Federation (JCF) and (2) "to develop a model or

set of models that would produce a standardized assessment of the

agencies and their associated services; (Mantel et al., 1972, pp. 1-2).

As a first step in defining the output of the JCF system,

answers to the following questions were sought;

1. What are the goals of the system?
2. What services are being delivered?
3. Who received the services?
4. Which agencies deliver the various service? (Ibid., p. 8)

General goals were first identified. Then answers to the last 3

questions enabled a 3 dimensional unit of account--the service-client-

agency package to be produced. The development of this unit of ac-

count permitted the "research task to be redefined as the development

of a measure of the 'output' associated with each service-client-

agency package" (Ibid., p. 11). Output was to be defined in terms

of (1) the number of clients served, (2) the amount of service time re-

ceived, (3) the value of the physical throughput (items 1 and 2), and (4) the quality of the physical throughput.

The physical throughput measures were obtained from agency records. The value measure, an indicant of the relative importance of each service-client-agency package, was considered "analogous to that of price in the economic measurement of industrial output" (Mantel et al., 1972, p. 13). The Delphi technique (see page 71 for a description of the technique) was used to develop value measures for each service-client-agency package. Four separate reference groups participated in the Delphi exercises which involved the assignment of 1 of 5 verbally described levels of importance to the service-client-agency packages. Whenever approximately 80% of each group's ratings were in any 2 contiguous categories of importance, agreement (within the group) as to the value of the service-client-agency package was deemed to exist (Reisman et al., 1970, p. 17).

In order to develop a measure of the quality of service the research team decomposed quality into 6 component criteria which were in turn decomposed into several elements. This rather complicated process involved

1. collection of element data

2. specification of weights (by a Delphi process) for combining elements into criteria

3. restatement of criteria scores as utility scores through a set of transformation functions relating individual criterion scores to their associated utility values (again a Delphi process

was used)

    4. specification of weights (via Delphi) for combining the criteria into a quality score.

Concerning this approach to developing a measure of quality, the research team noted that "since no 'true' measure of quality is known, the purpose here was to develop _indicators_ of those aspects of quality, the criteria, which appear to be important parts of the undefined quality whole" (Mantel _et al_., 1972, p. 16).

    The research team believed that the measurement model developed could be used to

    1. compare output data by service-client-agency package across time

    2. compare output data for a given service-client package between agencies offering the same package

    3. compare output data for different service-client packages either across time or between agencies.

Furthermore,

> ...actual use of the model focuses primarily on expected
> changes in budget allocation. Each operational change
> contemplated by an agency which has a change in the level
> of resources associated with it... can be evaluated in
> terms of the expected changes in the basic measures of system
> output... These expectations can be compared to those pre-
> sented by another agency seeking funds for its own proposals.
> The budget Committee, which must make decisions on how to
> allocate scarce resources, has access to quantified expec-
> tations for each of the competitors for resources. (Mantel
> _et al_., 1972, p. 30)

    Since the ultimate value of this or any other model for decision making must be ascertained thru its use, it is auspicious that the JCF Board of Trustees has committed the JCF to actual use of the

model for the next several years.

Although the pioneering work performed by the Case Western research team has served as an extremely useful guide for this research, certain important differences in approach and methodology exist between that work and this research. These will now be discussed.

In the JCF model convergence among rater groups was used as the criterion for assessing the validity of value measures. Concerning the process of validating subjective measures, Campbell and Fiske write

> For the justification of novel trait measures, for the valid-ation of test interpretation, or for the establishment of construct validity, discriminant validation as well as convergent validation is required. (1959, p. 31)

Both convergent and discriminant validity are being utilized in this research.

The quality measures generated in the JCF model are the result of a complex process involving the identification, measurement and weighting of criteria hypothesized to contribute to quality. In this research, no attempt will be made to identify quality elements and then aggregate them into a quality measure. Instead a global measure of quality will be employed. Global measures are much simpler to work with and existing evidence suggests that such global measures reasonably approximate those obtained thru a more extensive process (Lawler, 1967, p. 370).

In the JCF model no attempt was made to identify relationships between measures of input and output. Because such relationships are believed to be of value to decision makers, an attempt will be

made in this research to identify input-output relationships through the use of regression analysis.

Political Rationality

Each of the methods previously examined represents, to some extent, a formalized, rational approach to decision making in the NFP area. Their proponents believe that these methods, by making explicit the goals, alternatives, cost, benefits, performance criteria, etc., offer a considerable improvement over decisions based on intuition, hunch and past experience.

However, the value of the formal rational approach (first method) has been challenged by a school of thought associated primarily with Charles Lindblom (1959; 1965). Lindblom (1959) describes another approach (second method) which he believes is more realistic and better suited to decision making in the NFP area. This method of decision making is based primarily on collective wisdom and the past experience of decision makers. Entitled successive-limited comparisons, it is characterized by the following:

1. Selection of value goals and empirical analysis of the needed action are not distinct from one another but are closely intertwined.

2. Since means and ends are not distinct, means-end analysis is often inappropriate or limited.

3. The test of a "good" policy is typically that various analysts find themselves directly agreeing on a policy (without their agreeing that it is the most appropriate means to an agreed objective).

4. Analysis is drastically limited:
     i) Important possible outcomes are neglected.
    ii) Important alternative potential policies
        are neglected.
   iii) Important affected values are neglected.

5. A succession of comparisons greatly reduces or elimi-
nates reliance on theory. (Lindblom, 1959, p. 81)

Concerning these two approaches, Lindblom writes

> For complex problems, the first [formal, rational] of these
> two approaches is of course impossible. Although such an
> approach can be described, it cannot be practiced except
> for relatively simple problems and even then only in a some-
> what modified form. It assumes intellectual capacities and
> sources of information that men simply do not possess and
> it is even more absurd as an approach to policy when the time
> and money that can be allocated to a policy problem is limited,
> as is always the case. Of particular importance to public
> administrators is the fact that public agencies are in effect
> usually instructed not to practice the first method. That
> is to say, their prescribed functions and constraints-the
> politically or legally possible-restrict their attention to
> relatively few values and relatively few alternative policies
> among the countless alternatives that might be imagined. It
> is the second method that is practiced.
>
> Curiously, however, the literature of decision making, policy
> formulation, planning and public administration formalize the
> first approach rather than the second, leaving public adminis-
> trators who handle complex decisions in the position of prac-
> ticing what few preach. (Ibid., p. 80)

An analysis of the federal budgeting process by Wildavsky (1964) re-

vealed that federal resource allocations decisions follow an approach

quite similar to that suggested by Lindblom.* Wildavsky (1966) main-

tains that methods which seek to rationalize the decision making

process are generally deficient in that they ignore the political

---

*For business decisions, Cyert and March (1963) express a view of
the decision making process which is quite similar to Lindblom's.

costs of decisions. While a policy (program, project, etc.) may be optimal according to certain efficiency or effectiveness criteria (e.g., cost-benefit analysis) it may be politically non-optimal and irrational. Wildavsky states that

> ...exchange costs are incurred by a political leader when he needs the support of other people to get a policy adopted. He has to pay for this assistance by using up resources in the form of favors (patronage, log-rolling) or coercive moves (threats or acts to veto or remove from office). By supporting a policy and influencing others to do the same, a politician antagonizes some people and may suffer their retaliation. If these hostility costs mount, they may turn into re-election costs-actions that decrease his chances (or those of his friends) of being elected or reelected to office. Election costs, in turn, may become policy costs through inability to command the necessary formal powers to accomplish the desired policy objectives. (1966, p.

In light of the difficulties encountered by social scientists seeking to evaluate social programs (e.g., Rossi and Williams, 1972; Campbell, 1969) and the fate of the federal PPB (see p. 22), the Lindblom thesis, with its emphasis on the political nature of decisions, should not be ignored.

Summary

This review of performance measurement in the NFP area has revealed the existence of some sophisticated methods for assessing the efficiency and effectiveness with which resources are allocated towards desired ends. It also has revealed that most of these methods are not widely used. Furthermore, where attempts to apply such methods have been made, the results achieved have not been equal to expectations.

While the methods surveyed in this chapter warrant a belief in the efficacy and feasibility of performance measurement, problems of implementation indicate that progress will come slowly. Researchers, therefore, should avoid the oversell and reconcile themselves to piecemeal progress thru empirical applications.

CHAPTER III

DESIGN AND DEVELOPMENT OF A PERFORMANCE
MEASUREMENT MODEL FOR THE GAINESVILLE
RECREATION DEPARTMENT

Introduction

As indicated in Chapter II, the problems encountered in imple-
menting the formal methods for performance measurement present a
challenge to those who believe in the efficacy and feasibility of
performance measurement in the NFP area. In response to that challenge,
the main thrust of this research project was conceived to be the de-
velopment, with the close cooperation of a NFP organization, of an
operational performance measurement model.

Once the decision was made to attempt to design and develop a
performance measurement model for a NFP organization, the cooperation
of a specific NFP entity had to be secured.

The Gainesville Recreation Department

Selection of the Gainesville Recreation Department

Since the researcher had been working with the City of Gainesville,
Florida in the performance of audit and systems work, this City was
a logical first choice. However, a model for the entire City was
well beyond the resources of the researcher. Therefore, the researcher
decided to select a single City department to work with.

The Recreation Department was chosen because

1. its operational structure conformed to that of the JCF--
the social service measurement model developed for the JCF has served
as a guide for this research (see p. 29)

2. it was a politically low-key department and therefore the
occurrence of events which could abort the research appeared remote

3. performance measurement in the field of recreation is vir-
tually non-existent (e.g., Hatry and Dunn, 1971; Kraus and Curtis,
1973); the area thus appeared to offer the opportunity of challenging,
long-term research.

Having decided upon the Recreation Department, the Gainesville
City Manager was approached with the purpose of the research and a
request for cooperation. The City Manager endorsed the research and
directed the researcher to the GRD. A lengthy interview with the Di-
rector and Assistant Director of Recreation revealed enthusiastic sup-
port for the research project and a promise of cooperation.

Description of Gainesville Recreation Department

The GRD is comprised of an administrative division, a maintenance
division and 4 operating divisions--aquatics, athletics, centers and
playgrounds. The 4 operating divisions are directly responsible for
the various recreation programs provided to the Gainesville community.
The primary responsibilities of these operating divisions will now
be discussed.

Aquatics. The primary responsibility of this division is to
provide water recreation (primarily swimming pool). Its major areas

of activity are public swimming, instructional swimming and competitive swimming.

Athletics. The primary responsibility of this division is to provide organized team sports. Its major areas of activity include baseball, basketball, football, softball and instruction in various athletic skills.

Centers. This division provides indoor activities which include arts and crafts, dance, music and numerous games. The recreation centers also serve as a meeting place for numerous civic, educational and social organizations.

Playgrounds. This division provides supervised social activities for youth throughout the City. Activities include team sports, arts and crafts, dramatics, gymnastics and field trips.

In addition to the activities sponsored by the operating divisions, the GRD is responsible for maintaining park and picnic facilities, tennis courts and racketball courts.

As shown in the organization chart for the GRD (Figure 1), the Director of Recreation reports to the City Manager (administrative head of City government) who in turn reports to the City Commission--the elected representatives of Gainesville residents. The City Commission appoints interested Gainesville residents to serve on Public Recreation Advisory Board (PRAB). This board advises the City Manager in the area of recreation and meets monthly with the Director of Recreation for the purpose of discussing recreation policies, programs and problems.
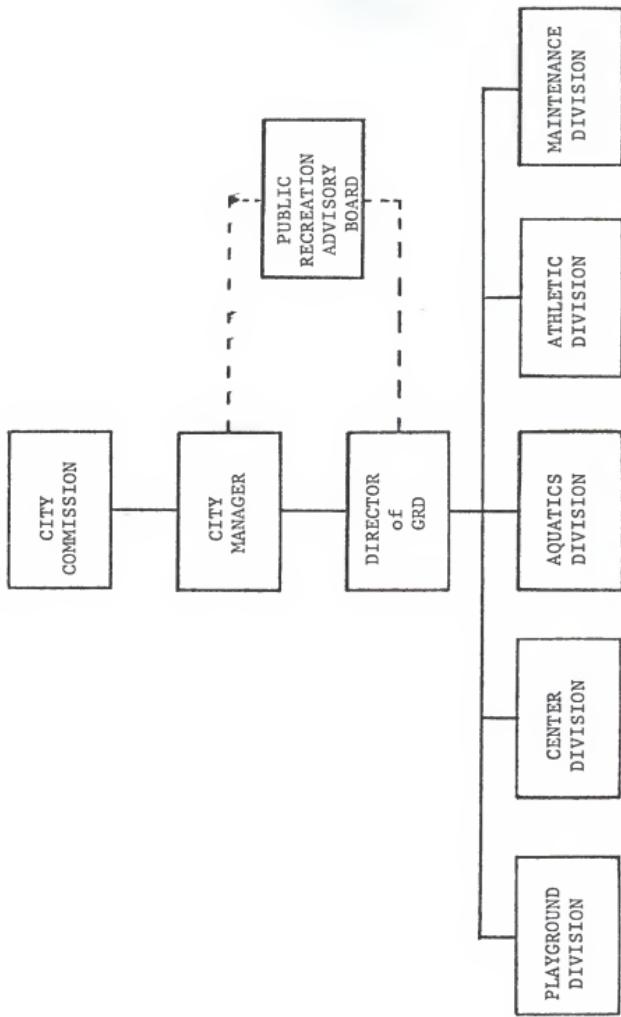
Figure 1. Organization Chart for the Gainesville Recreation Department.

Each month the GRD submits a report of its activities to the City Manager, City Commission and the PRAB. The report describes major recreation events occurring during the month and provides information on attendance at recreation facilities.

The GRD, along with all other City departments, prepares its budget for submission to the City Manager's Office in the spring. After numerous bargaining sessions, a final revised budget is prepared. It is submitted by the City Manager to the City Commission. After revisions desired by the City Commission have been made, the final budget is approved.

Recreation expenditures are controlled by this annual budget which reflects appropriations by division and by object of expenditure. Program budgeting is not used. For 1974-75, $465,087 was appropriated in the budget for recreation expenditures. Actual expenditures for 1973-74 were $384,407.

The GRD receives monthly financial reports which follow the budget format and depict (1) appropriations, (2) year-to-date expenditures and encumbrances, (3) unencumbered appropriations, and (4) monthly expenditures.

Status of Performance Measurement in the Gainesville Recreation Department

Information obtained through interviews with supervisors in the GRD and from an examination of departmental documents indicated that a formal system for evaluating department performance did not exist. Data by program (cost, number of participants, number of

staff, etc.) was not regularly collected and maintained. The monthly reports (see p. 41) submitted to the City Manager, City Commission and PRAB are unsuited for decision making. (Based on the extensive number and size of errors found in these reports and on conversations with GRD supervisors, it became evident that the reports were not being used for decision making.)

Departmental objectives were extremely vague. Objectives by division and program had not been defined.

In summary, performance measurement is at best based on ad hoc intuitive evaluations.

The preceding remarks are not intended as an indictment of the GRD. Rather they are made for the purpose of (1) indicating the undeveloped state of performance measurement and (2) illustrating the amount and type of preliminary work which must be done before a formalized system of performance measurement can be developed. The situation in the GRD is characteristic of much of the NFP area.

The model developed in this research should be evaluated in light of the present state of performance measurement in the NFP area. It should not be viewed as an extension of an already sophisticated decision making process.

The Performance-Measurement Model

Once the researcher had become familiar with the organizational structure and activities of the GRD, the design of a performance measurement model was begun. The model proposed is the result of the researcher's study of performance measurement methodologies (see Chapter II) and his knowledge of GRD activities. The model is

illustrated by means of Figure 2 and is discussed below. The con-
ceptual model (Part I of Figure 2) is presented first. The proced-
ures taken to operationalize the conceptual model (Part II of Figure
2) are then presented. (The numbers in brackets in the following
discussion refer to the relevant facets of Figure 2.)

Conceptual Model

For the purposes of this research, the GRD is viewed as a
social system whose purpose (assumed) is to promote the welfare
of the Gainesville community thru the provision of recreation pro-
grams and facilities. Ideally, the amount and type of public recre-
ation provided would be the amount and type consistent with the
maximum total welfare for the Gainesville community. Assuming
knowledge of the social welfare function for the community, production
functions, and cost functions, the conditions for maximum welfare
can be identified and the amount and type of public recreation
required for a welfare maximum can be determined. A mathematical
example of this process follows:

Let

P_i $i = 1,\ldots,n$
      be recreation programs (outputs)

H_j $j = 1,\ldots,w$
      be inputs

$P_i = P_i$ $(H_{i1},\ldots H_{iw})$
      be production functions; H_{ij} is the amount of the j th.
      input used in producing the i th. program

$C_i = C_i$ $(H_{i1},\ldots,H_{iw})$
      be the cost of program i

Figure 2. Performance Measurement Model for the Gainesville
Recreation Department.
Part I—Conceptual Model--above solid line.
Part II—Operational Model--below solid line.
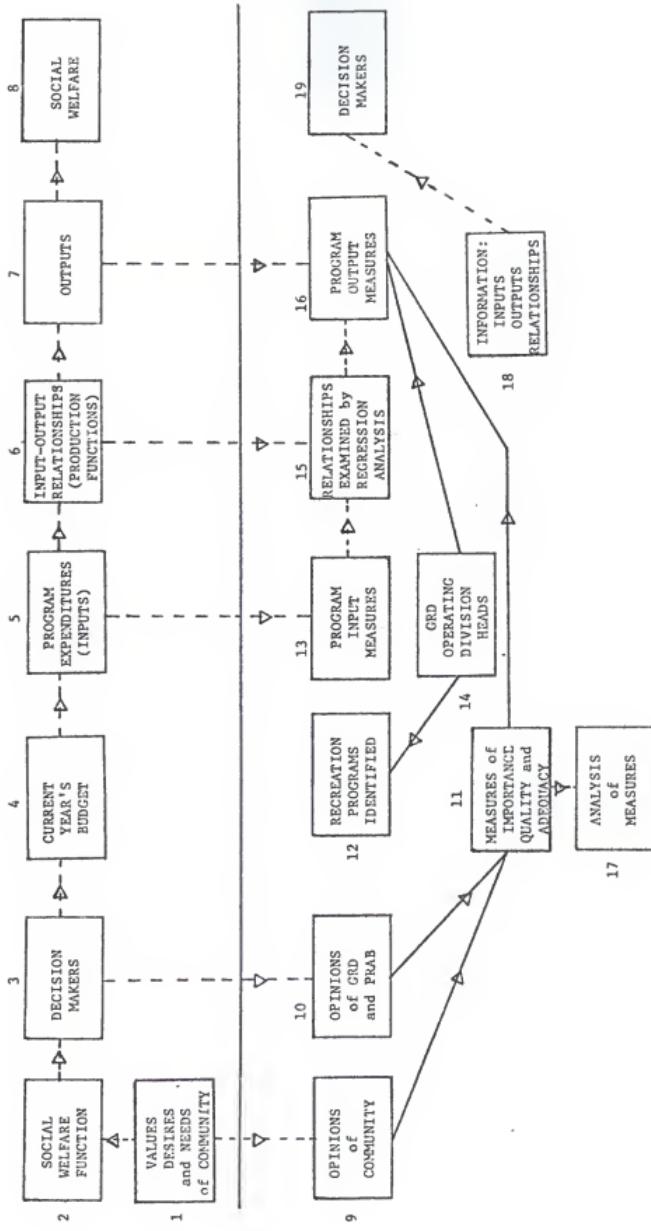
$W = W(P_1,\ldots,P_n,C_1,\ldots,C_n)$ be the social welfare function; $\frac{\partial W}{\partial P_i}>0$; As P_i is increased

welfare increases; $\frac{\partial W}{\partial C_i}<0$; Increasing P_i entails reducing

the output of some other good and thereby the welfare derived from its consumption--this welfare loss is the "true cost" of producing one more unit of P_i

W can be maximized subject to the production and cost functions.

Substituting the production functions for P_i and cost functions for

C_i yields $W = W[P_1(H_{11},\ldots,H_{1w}),\ldots,P_n(H_{n1},\ldots,H_{nw}),C_1(H_{11},\ldots H_{1w}),\ldots$

$$C_n(H_{n1},\ldots H_{nw})]$$

Calculate the partials of W with respect to H_{ij} and set them equal

to zero: $\frac{\partial W}{\partial H_{ij}} = W_{Pi}\frac{\partial P_i}{\partial H_{ij}} + W_{Ci}\frac{\partial C_i}{\partial H_{ij}} = 0$

The first term of each partial represents the gain in welfare from

the use of 1 more unit of H_j in the production of P_i. The second

term represents the loss in welfare from such usage. Since $W_{Ci}<0$,

the sum of the 2 terms for a partial may be interpreted as the

net welfare gain (loss) from the production of one more unit of P_i.

Solving the set of partial equations for all H_{ij}s would permit the

determination of all P_i and C_i. Given C_i, the optimal recreation

budget is $\sum_{i=1}^{n} C_i$.

As depicted in Figure 2, Part I, the values, desires and needs

of the members of the Gainesville community [1] enter into the de-

termination of the social welfare function [2]. Through

knowledge of this function, decision makers [3] can establish the optimal recreation budget [4]. With this budget the necessary program inputs [5] can be acquired. These inputs are transformed via the production functions [6] into those outputs [7] consistent with maximum welfare [8].

Operational Model

Limitations and plans

Since $W = W(P_1, \ldots, P_n, C_1, \ldots, C_n)$ is unknown, the amount of each P_i consistent with a welfare maximum cannot be determined. Furthermore, while benefits expressed in dollars (a cost-benefit approach) can sometimes be used as a surrogate for utility (Cohn, 1972), the dollar value of the benefits of the GRD recreation programs cannot be reliably determined (see p. 17). Thus at the operational level, the amount and type of public recreation to be provided is outside the purview of the model and must be left to the judgement of decision makers.

While the social welfare function (analogous to the profit function in the profit sector) appears beyond current knowledge, the other elements involved in the welfare maximization process do appear susceptible to identification, measurement and analysis. In the profit sector, for example, outputs, inputs, production functions and cost functions are an integral and routine part of management's information system. Such information should be of considerable value to NFP decision makers for it would reduce some of the uncertainty associated with the welfare maximizing process

(see p. 43) and thereby render this process both more efficient and effective.

The purpose of the operational model for the GRD (Figure 2, Part II) is to develop information [18] about program outputs, inputs, (including costs) and production for the use of Gainesville decision makers [19]. Such information should be useful in assessing the contribution of individual recreation programs to the welfare of the Gainesville community and in evaluating the impact of budget changes on outputs and costs.

Program identification

The first action taken was to identify the major recreation programs offered by the GRD. Once identified these programs served as the basis for producing measures of input and output and the determination of relationships between inputs and outputs. Each of the four operating division heads [14] was asked to supply the researcher with a list of major recreation programs [12] provided by his division during the fiscal year ending September 3, 1974.* These programs were reviewed and revised by the Assistant Director of Recreation. The revised list of programs was then reviewed by the division heads who concluded that the programs on the list were indicative of major recreation activities. Furthermore these programs were acceptable to them for use as the basic unit for obtaining

---

*Since the researcher began working with the GRD in February, 1975, fiscal 1974 represented the most recent and complete year of operations.

input and output measures.

In total 55 programs (activities) were identified. These programs, by division, are listed in Appendix A.

Output measures

While output measures for recreation programs have been largely ignored by recreation administrators (Hatry and Dunn, 1971, p. 14), the use and value of measures of the quantity and quality of recreation output have been discussed in the literature (Mack and Myers, 1965; Kraus and Curtis, 1973; Hatry and Dunn, 1971; Wennergren and Fullerton, 1972). In the social service measurement model developed for the Jewish Community Federation, the output of each type of service was defined in terms of its quantity, quality and importance (see p. 29). Since a GRD recreation program is analogous to a "type of service", similar measures would appear to be useful in defining recreation output.

The potential usefulness of measures of the quantity, quality and importance of recreation programs was discussed with the Director and Assistant Director of the GRD, division heads and members of the PRAB. All groups expressed the opinion that such measures would be useful.

Several PRAB members also stated that information concerning the adequacy of recreation facilities would be useful. Therefore a measure of facility adequacy will be included among the output measures (see p. 64).

Quantity of output

The number of <u>direct</u> user-service hours provided by a recreation program will be used as a measure of <u>output quantity</u> [16]. Direct users are primarily program participants. However, some programs also involve spectators (e.g., ball games, swim meets, etc.). User-service hours will be obtained by multiplying the number of participants and spectators by their respective average number of hours of usage.

While it may be appealing to add participant and spectator hours together for a single measure of physical output, such aggregation would represent an arbitrary assignment of equal weights to each type of output quantity. Therefore each type will be reported separately so that decision makers can assign whatever weights they may believe appropriate.

Because records of the number of program users and hours of usage are not maintained by the GRD, it will be necessary to rely on estimates provided by division heads in the GRD. This use of estimates, while necessary, introduces the potential for error—results of analyses based on such estimates must be interpreted cautiously with this limitation in mind.

As measures of output quantity, by themselves, provide insufficient evidence of how well the community is served by a program, measures of program quality and program importance will also be obtained.

Quality and importance of output

Quality. Program quality, as used in this research, means "how good the program is." While the determinants of program quality have not been identified and verified empirically, most of these determinants are believed to be under the control of and therefore the responsibility of the GRD.

Importance. Program importance, as used in this research, means "how much a program contributes to making the Gainesville community a more enjoyable place to live." This measure is presumed to reflect the values and preferences of the Gainesville community, to be fairly stable over time and to be independent of GRD activities. Measures of program importance generated by the GRD can be compared with measures generated by the PRAB and the community in order to obtain evidence as to whether or not the GRD is correctly assessing the values of the community.

While direct and objective measures of the quality and importance of recreation programs do not exist, it is possible to use opinions of knowledgeable people to produce indicants [11] of the quality and importance of recreation programs (Helmer, 1966; Dalkey et al., 1972; Mantel et al., 1972). Opinions for this purpose will be obtained by means of a questionnaire (see p. 59) which will be given to randomly selected households [9] in the Gainesville community, all supervisors in the GRD [10] and PRAB members [10] (see p. 57). The validity of these indicants of quality and importance will be assessed [17] by use of a multitrait-multirater methodology

Input measures

While the measures of program input pose no operational diffi-
culty, per se, the absence of cost accounting and program budgeting
means that desired measures will have to be produced specifically
for this research. The operating division heads will attempt to
provide the following data for each of their programs (out of the
55 identified at p. 49) for the year ended September 30, 1974:

1. Direct labor cost

2. Other direct cost

3. Number of staff and staff hours

4. Number of volunteers and volunteer hours

5. User fees

Items 1 and 2 above refer to the variable delivery cost of
the program. Direct labor accounts for approximately 70% of total
direct costs. No attempt will be made to obtain information on
fixed costs--these costs are not readily available and their allo-
cation to individual programs would be arbitrary.

Items 3 and 4 refer to the labor (paid and unpaid) used di-
rectly in providing a program. While data on other inputs would
also be desirable, it was not available.

While many programs are offered free of charge, some require
payment of entry, usage or materials fees (item 5 above). Such fees
represent a reduction in the cost of the program to the community

at large.

The absence of formal information systems for the routine collection of these input measures means that estimates will again have to be used for certain items in certain programs. Therefore the caveat issued in connection with user-service hours (see p. 50) is applicable here.

Input-output relationships

Once the input and output measures have been obtained, regression analysis (see p. 75) will be used in an attempt to identify relationships between labor inputs and the quantity and quality of output* [15]. While such relationships are analogous to production functions, the input and output measures to be used refer to different programs instead of different levels of input and output for the same program.

In its production of a particular recreation program, it appears reasonable to assume that the GRD intends to produce a certain quantity and quality of output. Furthermore, it is probably possible to tradeoff quantity and quality in the production of output with the result that a given output can be produced with various combinations of quality and quantity. This tradeoff is depicted graphically in Figure 3.

---

*Since the GRD has no control over the importance of programs, measures of importance are excluded from this analysis. Since the labor input measures to be collected relate only to participant hours, spectator hours are excluded from user hours for this analysis.
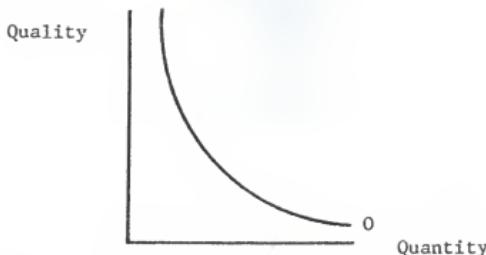
Figure 3. Tradeoff between quantity and quality in the production of recreation output.

If the nature of the quantity–quality interaction were known, it would be possible to aggregate the quality and quantity measures produced in this research into a scalar index of output which could then be related to inputs for the purpose of identifying recreation production functions. Since the nature of this interaction is not known, the relationships to be examined will be limited to the following:

G1. $U = U(L)$

G2. $U_e = U_e(L)$; $Q = Q^e$; $e = 1, 2, 3$

G3. $Q = Q(\frac{L}{U})$

where U = user hours

L = labor hours

Q = quality

Q^e = low, average and high quality respectively

For G1 the relationship between output quantity and labor inputs is to be examined for all programs without regard to program quality. For G2 the observed program quality range will be divided into 3

categories (low, average and high quality) and the relationship between output quantity and labor inputs for the programs in each quality category will be examined. For both G1 and G2 it is expected that $\frac{dU}{dL} > 0$. For G3 the relationship between quality and the labor-user ratio will be examined. It appears reasonable to assume that

$$\frac{dQ}{d\left(\frac{L}{U}\right)} > 0.$$

If the relationships for G1, G2 and G3 can be described successfully, decision makers should be better able to predict the effects of input changes on program quantity and quality. Furthermore, knowledge of G2 may provide insight into the nature of the quantity-quality tradeoff in the production of program output.

Simple linear models will first be applied to G1, G2 and G3. If such models are found to be inadequate, more complicated models will be utilized.

Uses of the Performance Measurement Model

While the operational performance measurement model generates quantitative information which can be used for many different purposes at different levels of decision making, the most important uses appear to be the following:

1. an assessment of the degree to which the GRD and PRAB correctly perceive the values, desires and needs of the Gainesville community. A prerequisite for optimal performance (from the community's point of view) would appear to be agreement on program importance between decision makers and community members. Furthermore, unless

there is agreement between decision makers and community members about which recreation facilities are adequate or inadequate, capital budgeting decisions will probably be unsound.

2. a comparison of recreation programs on the basis of (1) physical output, (2) quality of output, (3) importance of output, and (4) cost of output. These comparisons can be made between different programs within a particular year or over a number of years and between the same program over a number of years. These comparisons should improve resource allocation decisions and will provide direction for corrective action.

3. an evaluation, based on quantitative information (as noted in 2 above), of changes in GRD performance over time.

CHAPTER IV

METHODOLOGY

Introduction

The model and the procedures involved in operationalizing it were
set forth in the preceding chapter. In this chapter certain facets
of the model and the methodologies used in its implementation will
be discussed in more detail. To be discussed are (1) the rater groups
used to produce the indicants of program importance and quality and
facility adequacy; (2) the development of the questionnaire used to
obtain opinions from the rater groups; (3) the multitrait—multimethod
methodology; (4) the Delphi technique; and (5) regression analysis.

Rater Groups

Once the decision to use opinions to produce indicants of pro-
gram quality and importance had been made, it was necessary to decide
whose opinions were to be used. Concern (1) for assessing the validity
of indicants produced, (2) a desire to determine how well the decision
makers (GRD and PRAB) reflected the views of the community they served
and (3) a desire to obtain citizen input resulted in the choice of the
following 3 independent groups: GRD supervisors; PRAB members; com-
munity members. These groups will now be discussed.

GRD Supervisors

All 14 supervisors in the GRD agreed to participate. Since these

supervisors are primarily responsible for initiating, maintaining and eliminating recreation programs, they represent a very knowledgeable group in regard to recreation programs. Because there is a large amount of interaction and cooperation among the supervisors, their knowledge is not limited to those programs for which they are directly responsible.

While their knowledge of recreation programs militates in favor of valid appraisals, their closeness to recreation activities and their vested interest in the GRD's performance could inject bias into their opinions (consciously and unconsciously).

PRAB Members

The members* of the PRAB also agreed to participate. Appointed to the PRAB because of their interest in and knowledge of recreation, board members, like the GRD supervisors, constitute a group of recreation experts. However, board members are independent of the GRD and have no direct vested interest in the GRD's performance.

The opinions of the GRD supervisors and PRAB members will be obtained by formal questionnaire within a Delphi (see p. 71) framework.

Members of the Gainesville Community

The members of the Gainesville community are the primary users and beneficiaries of recreation programs and facilities. Through their tax dollars they bear most of the costs of providing these programs and facilities. Therefore, they represent a valuable source of information

---

*All 9 official members and 2 ex-officio members participated.

in regard to the importance and quality of recreation programs and the adequacy of recreation facilities.

Ideally, the opinions of both program participants and a random sample of members of the Gainesville community should have been obtained. Unfortunately, for most recreation programs, participants were not known and therefore direct contact with them was impossible. However, this limitation could be somewhat overcome by developing a survey instrument which distinguished between participants and non-participants.

Primarily because of resource constraints the opinions of a random sample of community members will be selected by mail survey.

While an up-to-date list of individual members of the Gainesville community was not available, a list of households was available in the form of the City's Utility Billing Listing. (This listing contains the name and address of all the customers of City electric, water and sewer service). Since the City is the sole source of electricity to the community, this list encompasses most of the households and therefore most of the members of the community.* Furthermore, the list is current and accurate.

From this Utility Billing Listing of approximately 33,000 residental accounts 2,000 were selected at random for the mail survey.

Questionnaire Development

The process of designing and testing a questionnaire for use

---

*The billing listing obviously does not encompass all members of the community and therefore the information obtained through its use will be somewhat biased. This is a limitation of the research.

in obtaining the opinions of the rater groups proved to be quite extensive. The stages of questionnaire development will now be discussed.

The initial questionnaire designed by the researcher was reviewed by the members of the dissertation committee, GRD supervisors and colleagues of the researcher. After necessary revisions were made, it was presented to PRAB members at their monthly meeting with the Director of Recreation. After incorporating certain changes suggested by PRAB members, the questionnaire and a questionnaire evaluation form were given to all 34 employees of the GRD for completion and evaluation. Based on the results of this pretest, the questionnaire was revised for the final time. (See Appendix B for the final questionnaire.)

Because control over responses of households selected for the mail survey was limited to the questionnaire itself, the questionnaire was designed specifically for this target group. The questionnaire so produced was easily modified for use with the GRD and PRAB.

In final form the questionnaire (Appendix B) used in the mail survey consisted of a cover letter and 5 parts:

Part 1	General Information
Part 2	Importance of Programs
Part 3	Quality of Programs
Part 4	Participation in Programs
Part 5	Adequacy of Facilities and Programs

The most important facets of this questionnaire will now be discussed.

Cover Letter

The cover letter contains

1. an appeal from the Director of the GRD for the cooperation of community members. This appeal directly associates the questionnaire with the GRD and should increase the response rate.

2. operational definitions of importance and quality: "By importance, we mean how much a program contributes to making the Gainesville community a more enjoyable place to live"; "By quality we mean how good the program is." These definitions should help insure that respondents are replying to the same questions.

3. instructions for completing the questionnaire.

Socio-Economic Data

In order to encourage community members to express their opinions, their anonymity was guaranteed. Such anonymity, however, precluded a followup of non-responses and thereby introduced the possibility of self-selection bias--the opinions of those returning the questionnaire may not be representative of those not returning the questionnaire. This possibility imposes limitations on the conclusions which can be drawn from the results of the survey.

In an effort to identify the existence of self-selection biases, the following socio-economic information was requested of respondents: sex; age; education; marital status; number living in home; family income; college status; and reside inside or outside City limits. This survey information will be compared with that for the entire community (based on 1970 census tract data) in an effort to assess the representativeness of the survey.

Indicants of Program Importance and Quality

The approach* followed in generating indicants of program im-

---

*This approach is based on that used for the JCF (see p. 30).

portance and quality will now be discussed.

Importance. Each group member was asked to select from 5 cat-
egories of importance the 1 category which best represented his opinion
of the importance of a recreation program to the Gainesville community.
The categories of importance and their numerical ratings are

Very low	1
Low	2
Average	3
High	4
Very high	5

Quality. Each group member was asked to select from 5 categories
of quality the 1 category which best represented his opinion of the
quality of the program. The categories of quality and their numerical
ratings are

Very poor	1
Poor	2
Fair	3
Good	4
Very good	5

Rather than force a respondent to express an opinion of the quality of
a program with which he was not familiar, the respondent was permitted
to select "no opinion" as his response.

So that the verbal categories can be described statistically,
numerals from 1 to 5 will be assigned to the categories. While tech-
nically speaking the measures so produced are ordinal, they will be
treated as approximately an interval scale. Such treatment is common
in social science (Kerlinger, 1973, pp. 439-441). The mean of the
numerical values assigned to the opinions of group members will con-
stitute the indicants of program importance and quality.

Questionnaire Length and Program Selection

The rate of response to mail questionnaires is normally inversely related to the length of the questionnaire (Parten, 1950, p. 385). Furthermore, long questionnaires are likely to be poorly filled out (Ibid.).

Therefore, the desideratum of as large and complete a response as possible appeared inconsistent with the inclusion of all 55 recreation programs in the questionnaire. (The pre-test questionnaire, which contained only 15 programs, required between 15 and 20 minutes for completion.) The procedure finally decided upon was to select at random 24 of the 55 programs and to then assign at random these 24 programs among 2 questionnaires (A and B) which were identical except for programs. Questionnaires A and B will be alternated among the randomly selected sample of 2,000 households (ABABAB...).

This method of alternately assigning questionnaires A and B to the households should produce comparable response groups. If the 2 groups are comparable, their response should be similar.

In order to determine the similarity of responses to questionnaires A and B, 4 of the same programs were included on each questionnaire. Indicants of program importance and quality for these 4 programs can be tested statistically for agreement between the 2 groups. The chi-square (X^2) test will be used for this purpose.

Order of Programs on Questionnaire

The order of recreation programs in questionnaire parts 2, 3, and 4 was randomly determined for each part. Thus, while programs are

the same across the 3 parts, their order differs. This procedure was followed in order to discourage conscious coordination by respondents of their responses across the 3 parts.

Participation in Programs

In questionnaire Part 4 respondents were asked to indicate the frequency of participation (by the respondent and his family) in the recreation programs. This information was solicited in order to

1. determine if opinions of program importance and quality are related to the frequency of participation in a program. These relationships will be analyzed by means of the X^2 test.

2. assess the validity of opinions of questionnaire respondents. It appears reasonable to assume that for most respondents knowledge of program quality comes from participation in the program. If this assumption is correct, "no opinion" responses for program quality should be associated with the participation category "never participate." The absence of such an association would appear to suggest that program quality was evaluated in a capricious manner and that therefore the opinions obtained are unreliable.

Adequacy of Recreation Facilities

During an examination of the initial questionnaire by the PRAB, several board members raised questions concerning the adequacy of recreation facilities and the impact of such facilities on the quality of recreation programs. This resulted in the addition to the questionnaire of Part 5 where respondents are requested to express their opinion as to the adequacy or inadequacy of recreation facilities.

Adequacy will be assigned the numeral 1 and inadequacy the numeral 2. Indicants of facility adequacy will be produced and the association between facility adequacy and the quality of certain programs will be examined by use of X^2.

The agreement of the adequacy ratings produced by the 3 groups will be examined by use of the Pearson product-moment correlation coefficient. Within the community group, the existence of significant correlations between respondents A and B would provide additional evidence for the existence of comparable response groups (see p. 63).

Geographic Location of Respondent

An identification of the general geographic location of household respondents was viewed as desirable because

1. recreation values, desires, needs and experiences may vary from location to location within the community

2. certain socio-economic characteristics can be identified with certain geographic locations (based on 1970 census tract data) and therefore the geographic location of respondents will be helpful in identifying response biases.

The format of the Utility Billing Listing suggested the means by which the identification of the general geographic location of respondents could be made. Customer accounts (households) are grouped into 6 billing cycles each of which can be identified with a particular geographic location in the Gainesville community:

Cycle 1   Predominantly southeast with some northeast

Cycle 2   Northeast

Cycle 3	Predominantly southwest
Cycle 4	Northwest
Cycle 5	Predominantly southwest with some northwest
Cycle 6	Northwest

By color-coding the questionnaires and sending questionnaires of the same color to the accounts within a cycle, an identification of the general geographic location of respondents can be made. This will permit (1) the generation of indicants of importance, quality and adequacy by geographic location and (2) the calculation of response percentages by geographic location.

Modification of Survey Questionnaire for Use with GRD and PRAB

The survey questionnaire required little modification for use in obtaining the opinions of the GRD and PRAB. The modifications were

1. elimination of the general information section (Part1). This information was not needed.

2. increase in number of programs evaluated (from 14 to 55).

3. addition of columns for feedback information (for Delphi). The same questionnaire (Appendix E) was used for GRD and PRAB members. The cover letters (Appendix F) differed only slightly from one another.

Multitrait-Multimethod Methodology

Indicants of program importance and quality will be produced from the opinions obtained from the 3 independent groups. The multitrait-multimethod matrix, developed by Campbell and Fiske (1959), will be used to assess the validity of these indicants. This methodology has come to be recognized as a powerful tool for ascertaining the validity

of constructs (e.g., Kerlinger, 1973, pp. 464-466; Runkel and McGrath,

1972, pp. 163-167).

Concerning this method Campbell and Fiske state that

1. Validation is typically underline convergent, a confirmation by independent measurement procedures. Independence of methods is a common denominator among the major types of validity (excepting content validity) insofar as they are to be distinguished from reliability.

2. For the justification of novel trait measures, for the validation of test interpretation, or for the establishment of construct validity, discriminant validation as well as convergent validation is required. Tests can be invalidated by too high correlations with other tests from which they were intended to differ.

3. Each test or task employed for measurement purposes is a trait-method unit, a union of a particular trait content with measurement procedures not specific to that content. The systematic variance among test scores can be due to responses to the measurement features as well as responses to the trait content.

4. In order to examine discriminant validity, and in order to estimate the relative contributions of trait and method variance, more than one trait as well as more than one method must be employed in the validation process. In many instances it will be convenient to achieve this through a multitrait-multimethod matrix. (1959, p. 81)

To illustrate their validation process they present a synthetic multi-

trait-multimethod matrix. (This matrix is included here as Table 1.)

In terms of this matrix, Campbell and Fiske state 4 criteria which

bear on the question of validity:

1. ...the entries in the validity diagonal should be significantly different from zero and sufficiently large to encourage further examination of validity....

2. ...a validity diagonal value should be higher than the values lying in the column and row in the heterotrait-heteromethod triangles....

TABLE 1

SYNTHETIC MULTITRAIT-MULTIMETHOD MATRIX

		Method 1			Method 2			Method 3		
		A1	B1	C1	A2	B2	C2	A3	B3	C3
Method 1	A1	(.89)								
	B1	.51	(.89)							
	C1	.38	.37	(.76)						
Method 2	A2	.57	.22	.09	(.93)					
	B2	.22	.57	.10	.68	(.94)				
	C2	.11	.11	.46	.59	.58	(.84)			
Method 3	A3	.56	.22	.11	.67	.42	.33	(.94)		
	B3	.23	.58	.12	.43	.66	.34	.67	(.92)	
	C3	.11	.11	.45	.34	.32	.58	.58	.60	(.85)

Source: Donald T. Campbell and Donald W. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, March, 1959, Table 1, p. 82, by permission.

Note: The validity diagonals are the three sets of underlined values. The reliability diagonals are the three sets of values in parentheses. Each heterotrait-monomethod triangle is enclosed by a solid line. Each heterotrait-heteromethod triangle is enclosed by a broken line.

3. ...a variable [should] correlate higher with an inde-
pendent effort to measure the same trait than with measures
designed to get at different traits which happen to employ
the same method....

4. ...the same pattern of trait interrelationship be shown
in all of the heterotrait triangles of both the monomethod
and heteromethod blocks. (1959, pp. 82-83)

The first criterion provides evidence of convergent validity; the next

3 criteria provide evidence of discriminant validity.

While Campbell and Fiske present their methodology in terms of

independent methods, it is evident from the examples they discuss that

independent raters can be viewed as different methods (1959, pp. 89-97).

The use and application of a multitrait-multirater approach for mea-

suring managerial performance has been discussed by Lawler who con-

cludes that "this approach has advantages for establishing criteria

where they are needed, either for research purposes or for personnel

decision-making purposes" (1967, p. 369).

As employed in this research the 3 independent rater groups

will constitute the different methods and indicants of quality and

importance of the traits. This combination of raters and indicants

will be hereafter referred to as the multivariable-multirater matrix.

An illustration of the multivariable-multirater matrix to

be used in this research is presented in Table 2. Correlation

between all 3 groups for the 24 programs commonly evaluated

will be obtained. Correlation between the GRD and PRAB for all 55

programs will be obtained. Thus 1 matrix will involve 3 rater

groups and 2 variables; another, 2 rater groups and 2 variables.

TABLE 2
MULTIVARIABLE-MULTIRATER MATRIX

		Opinions of Recreation Supervisors		Opinions of Advisory Board Members		Opinions of Community	
		Program Impt.	Program Quality	Program Impt.	Program Quality	Program Impt.	Program Quality
Opinions of Recreation Supervisors	Program Impt.	R	M	C	H	C	H
	Program Quality		R	H	C	H	C
Opinions of Advisory Board Members	Program Impt.			R	M	C	H
	Program Quality				R	H	C
Opinions of Program Users	Program Impt.					R	M
	Program Quality						R

Notes: Adapted from Runkel and McGrath, 1972.

R—reliability; C—convergent validity; M—rater variance; H—heterovariable-heterorater correlation. M and H correlations in relation to C correlations reflect discriminant validity.

The opinions of the GRD and PRAB will be obtained several times and the correlation between ratings based on opinions expressed at different times will provide measures of reliability. Since opinions of community members will be obtained but once, it will not be possible to assess the reliability of their ratings in a traditional sense. However, some insight into the stability of their ratings can be obtained from the correlation between common program ratings of the A and B groups.

The Delphi Technique

In an effort to produce valid and reliable indicants of importance and quality, the Delphi technique will be used in obtaining the opinions of GRD and PRAB members. The technique has been well described by Dalkey:

> The Delphi technique, is a method of eliciting and refining group judgements. The rationale for the procedures is primarily the age old adage "Two heads are better than one," when the issue is one where exact knowledge is not available. The procedures have three features: (1) Anonymous response—opinions of members of the group are obtained by formal questionnaire. (2) Iteration and controlled feedback—interaction is effected by a systematic exercise conducted in several iterations, with carefully controlled feedback between rounds. (3) Statistical group response—the group opinion is defined as an appropriate aggregate of individual opinions on the final round. These features are designed to minimize the biasing effects of dominant individuals, of irrelevant communications, and of group pressure toward conformity. (1969, p. v)

The technique was developed by the Rand Corporation for the purpose of forecasting future events (Helmer and Gordon, 1964). Experiments with almanac type data (Dalkey, 1969) and with short-

term forecasts (Campbell, 1966) have suggested that responses gen-
erated by a Delphi process are superior to those produced by individuals
or committees.

The extension of Delphi to the area of value judgements is quite
recent. In this regard Dalkey et al. state

> Most of the experiments which have been conducted to date
> have dealt with factual material. However, in some ap-
> plications, the procedures have been employed to deal
> with a quite different sort of material, namely, value
> judgements. Typical is the use of Delphi procedures to
> identify and rate the objectives of industrial enterprises
> or to assess the relative importance of military missions.
> From the standpoint of the decision maker, opinions about
> values and objectives are just as relevant to decisions
> as factual opinions about consequences. Hence, the question
> whether Delphi procedures demonstrate advantages with value
> material of the same sort as those for factual material
> is a question of direct importance. (1972, p. 55)

In considering the logic of the extension of Delphi to value judge-
ments, the authors point out that "...if a group of equally competent
individuals expresses a range of opinions concerning a value question,
then the average opinion is more likely to approximate the correct
answer then an individual judgement, given the presumption that there
is a correct answer to the value question" (1972, p. 56). Since there
is no presently known way of assessing the excellence of value judge-
ments, Dalkey et al. present 3 conditions, found to exist in Delphi
experiments with factual data, as a means of partially evaluating
the usefulness of the Delphi technique in producing a group judgement
on a value question:

> 1. <u>Reasonable distribution.</u> If the distribution of group
> responses on a given numerical value judgement is flat,
> indicating group indifference, or if it is U-shaped, in-
> dicating either that the question is being interpreted

differently by two subgroups, or there is an actual dif-
ference of assessment by two subgroups, then it seems
inappropriate to assert that the group considered as a
unit has a judgement on that question.

2. Group reliability. Given two similar groups (e.g.,
two groups selected out of a larger group at random) the
group judgements on a given value question should be sim-
ilar. Over a set of such value judgements, the corre-
lation for the two subgroups should be high.

3. Change and convergence on iteration with feedback.
This condition is proposed in part by analogy with results
from experiments with factual material, that is shifts of
individual responses toward the group response and reduction
in group variability. More generally, if members of the
group do not utilize information in reports of the group
response on earlier rounds when generating responses on
later rounds, it seems inappropriate to consider these
responses as judgements. (1972, p. 57)

These criteria were used by the authors to evaluate judgements generated

in a series of experiments dealing with the objectives of higher ed-

ucation and of everyday life and with the relative importance of these

objectives. Concerning the outcome of these experiments Dalkey et

al. state that "the results of applying the three criteria... to the

ratings of the educational and quality of life factors are all favorable

to the hypothesis that Delphi procedures are appropriate for formulating

group value judgements" (1972, p. 80). Subsequent to these experiments,

Dalkey et al. employed Delphi to develop and test an index of the "quality

of life" (1972, pp. 109-129).

Mantel et al. (1972) used a modified verson of the Delphi tech-

nique to develop measures of the value and quality of services provided

by the JCF (see p. 30, Chapter II). The researchers were pleased with

the results achieved; furthermore, the measures of quality and value

generated by Delphi were acceptable to JCF administrators.

The criteria proposed by Dalkey et al. and the degree of con-
sensus among the members of each group will be considered in the eval-
uation of the existence of a group judgement as to recreation program
importance and quality.

A decision to use Delphi entails consideration of the type of
feedback information and the number of iterations. The 2 most com-
mon forms of feedback which have been used in Delphi exercises are

1. Statistical feedback—generally the median group response and
the interquartile range from the previous round. Virtually all Delphi
studies have used this type of feedback.

2. Verbal feedback—this generally entails justification of
extreme responses (those outside the interquartile range). In some
Delphi experiments, participants, whose responses on the second round
are outside the first round's interquartile range, are asked to justify
their (relatively) extreme position. These justifications are then
fedback (along with statistical feedback) to all respondents in round
3. Round 3 respondents are permitted to provide counterarguments to
the justifications for extreme positions. The counterarguments are
fedback on round 4.

To date, the evidence is inconclusive as to which type of feedback
is most appropriate. However, verbal type feedback poses considerable
difficulties unless the number of participants and questions is extremely
small:

> If one includes all comments from all participants, the
> volume of feedback rapidly becomes prohibitive and its
> function self-defeating. On the other hand, the editing
> of first-round (or any round) data must necessarily be

somewhat arbitrary. When opinions are aggregated and con-
densed, certain participants will inevitably (and sometimes
justifiably) feel that their opinion has not been adequately
represented in the edited feedback version. The question
of validity also poses a difficult problem. (Thompson,
1973, p. 5)

Because of the number of participants (14 and 11) and questions
(174) involved in the Delphi exercises in this research, verbal type
feedback would be extremely difficult, awkward and time consuming.
Therefore it will not be used and feedback will consist solely of the
median group response and interquartile range on the previous round.

Iterations in previous Delphi studies have ranged from 1 to 6
and evidence as to the optimal number of the iterations is inconclusive.
A priori the number of iterations cannot be determined. Rather the
number of iterations will depend on (1) the degree of group consensus,
(2) amount of convergence between rounds and (3) the receptivity of
group members to participate in additional iterations.

Regression Analysis

The functional relationships hypothesized for the input-output
measures (see p. 54) will first be examined by using simple linear
regression analysis. The regression programs in Statistical Package
for the Social Sciences (SPSS) (Nie et al., 1975) will be employed for
this purpose. The strength of total relationships will be assessed
by examining r^2 and the regression coefficients will be tested for
significance with an F-test.

Plots of the data and residuals will be examined in order to as-
sess the validity of the linearity assumption. If the plots suggest
a departure from linearity, non-linear regression models will be applied

to the data using the SPSS polynominal regression program.

The regressions may reveal that the posited relationships are insignificant. While this may be due to the actual absence of re- lationships, it could also be due to the nature of the data employed:

1. The indicants of quality are based on opinions. While great care is being exercised in obtaining and evaluating these opinions, no conclusion as to their ultimate validity is warranted.

2. Participant hours, staff hours and volunteer hours are the results of estimates by division heads in the GRD. Thus relationships actually existing may be obscured by the unreliability of the data.

Statistical Procedures

The statistics used in this research can be found in Blalock (1972). Most of the statistical computations were performed by using the programs in SPSS. Unless indicated otherwise the .05 level of significance is being used.

CHAPTER V

DATA COLLECTION AND ANALYSIS

Introduction

This chapter is concerned with the collection and analysis of
the data specified by the operational model presented in Chapter III.
The methods used to collect and analyze this data were discussed in
Chapter IV. The data collection and analysis will be discussed in the
following order: (1) community survey data; (2) Delphi; (3) multivari-
able-multirater analysis; (4) facility adequacy; (5) validity of re-
liance on self-evaluations; and (6) input-output relationships.

Community Survey Data

Collection

The final questionnaire revisions were completed by May 1, 1975.
A cover letter (Appendix C) from the researcher to questionnaire re-
cipients and a self-addressed, postage paid return envelope were then
prepared. After the questionnaires (type A and B), cover letters and
return envelopes had been printed, they were placed in envelopes and
mailed, on May 28, 1975, to each of the 2,000 households in the sample.

Different colored questionnaires were used and the households
in each cycle were sent a different color of questionnaire. This made
possible the identification of the general geographic location of re-
spondents. Questionnaires A and B were alternated within each cycle

(ABABAB...).

The first responses were received May 30 and within 2 weeks 90%
of the total number returned had been received. By August 12, 486
questionnaires, representing a response rate of 24.3%, had been received.
The number of questionnaires mailed and returned are listed by cycle
in Table 3. In view of the length of the questionnaire, the response
rate was better than expected. However, it is still too low to permit
generalizing the results of the survey to the population from which
it was selected. This inability to generalize is a limitation of this
research.

Analyses

Comparison of response rates between A and B

Because A and B were identical except for programs (which were
randomly assigned to each) and because of the manner in which they
were distributed to households, no significant difference in response
rates was expected between A and B.

A and B response rates were compared by geographic location and
in total with a x^2 test. The results of the 2 tests are reported in
Table 4. The hypothesis that there was no difference in response rate
could not be rejected.

Comparison of response rates between geographic locations

One of the reasons for obtaining the geographic location of re-
spondents was to help in identifying response biases. Prior knowledge
of (1) the effect of socioeconomic characteristics on response rates
(Parten, 1950, p. 391) and (2) differences in socioeconomic characteristics

TABLE 3
ANALYSIS BY CYCLE OF QUESTIONNAIRES
MAILED AND RETURNED

	Question-naire	Number Mailed Out	Cycle as % of Total Mailed	Number Returned	Cycle as % of Total Returned	Number Returned as % of Number Mailed
Cycle 1 (Green) Predominantly Southeast with some Northeast	A	135		15		
	B	$\frac{134}{269}$	13.5%	$\frac{18}{33}$	6.8%	12.3%
Cycle 2 (Yellow) Northeast	A	118		22		
	B	$\frac{118}{236}$	11.8%	$\frac{19}{41}$	8.4%	17.4%
Cycle 3 (Pink) Predominantly Southwest	A	247		69		
	B	$\frac{248}{485}$	24.3%	$\frac{68}{137}$	28.2%	28.2%
Cycle 4 (White) Northwest	A	121		66		
	B	$\frac{121}{242}$	12.1%	$\frac{45}{111}$	22.8%	45.9%

TABLE 3—Continued

	Question-naire	Number Mailed Out	Cycle as % of Total Mailed	Number Returned	Cycle as % of Total Returned	Number Returned as % of Number Mailed
Cycle 5 (Blue) Predominantly Southwest with some Northwest	A	196		24		
	B	$\frac{196}{392}$	19.6%	$\frac{37}{61}$	12.6%	15.6%
Cycle 6 (Blue) Northwest	A	188		57		
	B	$\frac{188}{376}$	18.7%	$\frac{46}{103}$	21.2%	27.4%
Total All Cycles	A	$\overline{1000}$		$\overline{253}$		
	B	$\frac{1000}{2000}$	100.0%	$\frac{233}{486}$	100.0%	24.3%

TABLE 4

COMPARISON OF RESPONSE RATES

BETWEEN A AND B

Geographic Location	Frequencies				Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level
	A		B					
	Act.	Expt.	Act.	Expt.				
Predominantly Southeast with some Northeast	15	17	18	16				
Northeast	22	21	19	20				
Predominantly Southwest	69	71	68	66				
Northwest	66	58	45	53				
Predominantly Southwest with some Northwest	24	32	37	29				
Northwest	57	54	46	49	7.57	5	11.07	NS
	253		233					
In Total	253	243	233	243	.82	1	3.841	NS

*Blalock, 1972, p. 569. Unless indicated otherwise, significance level is .05.

**NS—significant at .05.

between locations in Gainesville (based on 1970 census) resulted in the
hypothesis that response rates would differ between locations. The
null hypothesis of no difference was tested with x^2 and <u>it was rejected</u>
(see Table 5). An examination of Table 3 reveals that Southeast, North-
east and Southwest with some Northwest (Cycle 5) are underrepresented
while the other locations, especially Northwest (white), are overre-
presented. Therefore, if the opinions of respondents also differ by
geographic location, the ratings of program importance and quality and
facility adequacy will be biased toward the overrepresented areas.

<u>Statistics produced</u>

The following statistics were computed for each of the 59 questions
appearing on the A and B questionnaires: (1) frequency distributions;
(2) histograms; (3) means; and (4) standard deviations. The mean impor-
tance and quality ratings are presented in Tables 19 and 20 (p. 115)
and the mean adequacy ratings are presented in Table 28 (p. 137).

<u>Comparability of A and B response groups</u>

Because of the methodologies employed in questionnaire preparation
and distribution (see pp. 63-64), the A and B response groups were
expected to be very comparable. Indeed, the existence of significant
differences in the responses of A and B to identical questions would
cast doubt on the stability and generalizability of the measures of
importance, quality and adequacy which were developed from the opinions
of survey respondents.

x^2 was used to test the hypothesis of no difference between A
and B in regard to

TABLE 5

COMPARISON OF RESPONSE RATES
BETWEEN GEOGRAPHIC LOCATIONS

	Frequencies		Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level
	Act.	Expt.				
Predominantly Southeast with some Northeast	33	66				
Northeast	41	57				
Predominantly Southwest	137	118				
Northwest	111	59				
Predominantly Southwest with some Northwest	61	95				
Northwest	103	91				
	486	486	83.7	5	20.517	.001

*Blalock, 1972, p. 569.

1. socioeconomic characteristics

2. ratings of importance, quality and participation for the
4 common programs (adult ceramics, springboard diving lessons, summer
track and tumbling lessons)*

3. ratings of facility adequacy.*

The comparisons made and related statistics are presented for the
above 3 categories in Tables 6, 7, and 8 respectively. For the 8
socioeconomic characteristics (Table 6), 4 program quality ratings
(Table 7) and 4 participation ratings (Table 7) the hypothesis that
there was no difference between A and B could not be rejected. Of
the 9 facilities evaluated (Table 8) as adequate or inadequate, for
only 1 (recreation centers) was the hypothesis of no difference re-
jected. For the 4 program importance ratings, however, the hypo-
thesis of no difference had to be rejected for each program (Table 7).

Although the preponderance of tests performed support the
hypothesis that A and B are comparable, the unanimous rejection
of this hypothesis for importance ratings was, at first quite puz-
zling. This problematic situation led to a search for a plausible
explanation.

A closer analysis of the programs contained on the A and B
questionnaires revealed that only A contained the 2 recreation programs

---

*Parametric tests were not used because the researcher was interested
in comparing the response distributions of A and B.

TABLE 6

COMPARISON OF A AND B RESPONSES--
SOCIOECONOMIC CHARACTERISTICS

Socioeconomic Characteristics	A Act.	A Expt.	B Act.	B Expt.	Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level**
Sex								
Male	170	172	161	159				
Female	82	80	71	73				
	252		232		.15	1	3.84	NS
Age								
16–20	12	15	16	13				
21–30	138	133	119	124				
31–50	56	62	63	57				
51–65	36	31	24	29				
Over 65	9	10	11	10				
	251		233		4.76	4	9.49	NS
Level of Education								
Grammar School	2		5					
High School	21	27	14	15				
Some College	65	68	65	62				
College Degree	165	158	145	152				
	253		229		2.56	2	5.99	NS
Married								
Yes	155	152	138	141				
No	94	97	93	90				
	249		231		.316	1	3.84	NS

TABLE 6—Continued

Socioeconomic Characteristics	Frequencies				Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level**
	A		B					
	Act.	Expt.	Act.	Expt.				
Number Living in Home								
1	42	42	38	38				
2	90	94	90	86				
3-5	112	107	93	98				
Over 5	6	7	8	15				
	250		230		1.12	3	7.82	NS
Family Income								
0-5,000	51	53	50	48				
5,001-10,000	41	48	51	44				
10,001-15,000	42	41	36	37				
15,001-20,000	37	38	37	36				
20,001-30,000	46	42	35	39				
Over 30,000	31	26	20	25				
	248		229		5.15	5	11.07	NS
College Student								
Yes	109	108	96	97				
No	140	141	128	127				
	249		224		.034	1	3.84	NS

TABLE 6—Continued

Socioeconomic Characteristics	Frequencies				Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level**
	A		B					
	Act.	Expt.	Act.	Expt.				
Live in Gainesville								
City Limit								
Inside City	170	171	156	155				
Outside City	80	79	72	73				
	250		228		.039	1	3.84	NS

*Blalock, 1972, p. 569. Significance level is .05.

**NS--not significant at .05.

TABLE 7

COMPARISON OF A AND B RESPONSES--
IMPORTANCE, QUALITY AND PARTICIPATION
FOR COMMON PROGRAMS

Common Programs	Frequencies A		Frequencies B		Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level**
	Act.	Expt.	Act.	Expt.				
Adult Ceramics								
Importance								
Very Low	27	24	19	22				
Low	50	40	27	37				
Average	111	98	79	92				
High	41	75	67	71				
Very High	8	75	30	71				
	237		222		28.08	4	18.47	.001
Quality								
Very Poor	5	4	2	3				
Poor	4	5	6	5				
Fair	13	16	18	15				
Good	25	23	19	21				
Very Good	16	19	20	17				
No Opinion	175	171	158	162				
	238		223		3.69	5	11.07	NS
Participation								
Never Partici- pate	216	218	204	202				
Sometimes Participate	27	30	22	28				
Frequently Participate	5		4					
	248		230		.314	2	5.99	NS

TABLE 7--Continued

Common Programs	Frequencies A		Frequencies B		Calculated X²	Degrees of Freedom	X² from Table*	Sig. Level**
	Act.	Expt.	Act.	Expt.				
Springboard Diving								
Lessons								
Importance								
Very Low	34	28	20	26				
Low	73	52	27	48				
Average	101	104	99	96				
High	24	40	52	36				
Very High	9	17	24	16				
	241		222		41.79	4	18.47	.001
Quality								
Very Poor	6	4	2	4				
Poor	2	3	3	2				
Fair	24	21	17	20				
Good	7	10	12	9				
Very Good	7	9	11	9				
No Opinion	192	191	178	179				
	238		223		6.51	5	11.07	NS
Participation								
Never Partici- pate	242	239	211	214				
Sometimes participate	4	10	9	8				
Frequently participate	3	2	2					
	249		222		2.105	2	5.99	NS

TABLE 7--Continued

Common Programs	Frequencies				Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level**
	A		B					
	Act.	Expt.	Act.	Expt.				
Summer Track								
Importance								
Very Low	18	18	16	16				
Low	49	40	27	36				
Average	116	106	87	97				
High	52	60	64	56				
Very High	8	19	29	18				
	243		223		21.16	4	18.47	.001
Quality								
Very Poor	4	4	3	3				
Poor	6	5	3	4				
Fair	21	21	20	20				
Good	21	24	25	22				
Very Good	9	13	17	13				
No Opinion	177	171	153	159				
	238		221		4.24	5	11.07	NS
Participation								
Never Partici-pate	218	216	200	202				
Somestimes Participate	24	30	24	29				
Frequently Participate	4		7					
	246		231		.310	2	5.99	NS

TABLE 7—Continued

Common Programs	Frequencies				Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level**
	A		B					
	Act.	Expt.	Act.	Expt.				
Tumbling Lessons								
Importance								
Very Low	40	30	18	28				
Low	64	53	38	49				
Average	98	103	99	94				
High	34	40	42	36				
Very High	6	17	26	15				
	242		223		29.21	4	18.47	.001
Quality								
Very Poor	6	6	5	5				
Poor	8	8	7	7				
Fair	15	16	15	14				
Good	16	19	20	17				
Very Good	5	8	11	8				
No Opinion	189	183	165	171				
	239		223		3.80	5	11.07	NS
Participation								
Never Partici-pate	231	229	214	216				
Sometimes participate	9	17	13	15				
Frequently participate	6		4					
	246		231		.538	2	5.99	NS

*Blalock, 1972. p. 569. Unless indicated otherwise, significance level is .05.

**NS—not significant at .05.

TABLE 8

COMPARISON OF A AND B RESPONSES--
ADEQUACY OF FACILITIES

Facilities	Frequencies				Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level**
	A		B					
	Act.	Expt.	Act.	Expt.				
Archery								
Adequate	125	132	129	122				
Inadequate	74	67	54	61				
No Opinion	54	54	50	50	2.30	2	5.99	NS
	253		233					
Baseball								
Adequate	175	173	158	160				
Inadequate	36	36	34	34				
No Opinion	42	44	41	39	.24	2	5.99	NS
	253		233					
Parks								
Adequate	111	104	88	95				
Inadequate	111	115	109	105				
No Opinion	31	34	36	33	1.81	2	5.99	NS
	253		233					
Playgrounds								
Adequate	116	119	112	109				
Inadequate	100	95	83	88				
No Opinion	37	39	38	36	.92	2	5.99	NS
	253		233					

TABLE 8--Continued

Facilities	Frequencies				Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level**
	A		B					
	Act.	Expt.	Act.	Expt.				
Racquetball & Handball								
Adequate	85	88	84	81				
Inadequate	130	127	113	116				
No Opinion	38	38	36	36	.36	2	5.99	NS
	253		233					
Recreation Centers								
Adequate	86	97	100	89				
Inadequate	132	117	92	107				
No Opinion	35	39	41	37	7.45	2	5.99	.05
	253		233					
Softball								
Adequate	175	176	166	165				
Inadequate	35	32	26	29				
No Opinion	43	45	41	39	.80	2	5.99	NS
	253		233					
Swimming Pools								
Adequate	71	77	76	70				
Inadequate	145	140	124	129				
No Opinion	37	36	33	34	1.41	2	5.99	NS
	253		233					

TABLE 8--Continued

Facilities	Frequencies					Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level**
	A		B						
	Act.	Expt.	Act.	Expt.					
Tennis									
Adequate	81	79	71	73					
Inadequate	139	139	128	128					
No Opinion	33	35	34	32					
	253		233			.34	2	5.99	NS

*Bialock, 1972, p. 569. Significance level is .05.

**NS--not significant at .05.

with the greatest public participation and visibility: park and
picnic facilities and public swimming. These 2 programs also re-
ceived the highest average importance ratings across the 3 groups
of raters. Now if the respondents had rated each recreation program
relative to the other programs contained in the same questionnaire,
the 4 common programs would not have received as high a rating by
the A group as by the B group (which did not contain any programs
rated "high" in importance). A comparison of the mean importance
ratings of the common programs between A and B (Table 9) revealed
that the means of B exceeded those of A and that the differences
were significant ($\underline{t}$ test) at the .05 level.

TABLE 9
COMPARISON OF A AND B MEAN IMPORTANCE RATINGS
FOR COMMON PROGRAMS

Program	Importance	
	A	B
Adult Ceramics	2.802	3.279
Springboard Diving Lessons	2.589	3.149
Summer Track	2.930	3.283
Tumbling Lessons	2.595	3.090

The impact of context on psychophysical measures is well
recognized (Parducci, 1974) and it appears that the different programs
appearing on the A and B questionnaires created contextual effects
which resulted in the observed significant difference in ratings
of program importance. While context should affect absolute ratings
of importance, it should not affect the rank order of the ratings.

Pearson's product-moment correlation coefficient for the common
program means is .9197 and is significant at .04. Thus based on
rank order, the importance ratings are comparable.

In summary, based on the results of tests performed, the
conclusion that A and B response groups are comparable appears
warranted.

Validity of survey responses

An evaluation of the validity of the opinions of survey respond-
ents was made in order to determine if these opinions could be used
in this research. The evaluation was based on both qualitative and
quantitative information.

Qualitative. Since the questionnaire was 6 pages long, con-
tained 59 questions and required about 20 minutes for completion,
it does not seem very likely that someone would complete it unless
he had a serious interest in it. Given such interest, legitimate,
non-capricious responses would be expected. Almost all of the question-
naires were completed in accordance with instructions--again suggesting
legitimate responses. Consistent with expectations, park and picnic
facilities and public swimming received high importance ratings--if
the questionnaires had been completed capriciously, such high ratings
would have been unlikely.

Quantitative. The quantitative assessment of validity is
based on the relationship between program participation and the ex-
pression of an opinion on program quality. Since it is reasonable
to assume that for most people knowledge of program quality comes

from participation in the program, those survey respondents who do
not participate in a program would be expected to express "no opinion"
as to quality. A test of the existence of such a relationship can
be conducted by use of the X^2 statistic. The null hypothesis is
that participation has no effect on the expression of an opinion
on quality. The contingency table used to test this hypothesis
is presented in Table 10 below. The test was conducted for each of
the 28 programs contained in questionnaires A and B. For all 28
programs the null hypothesis that participation has no effect was
rejected. For 27 programs the level of significance was .0001
while for the remaining program it was .004.

Based on the qualitative and quantitative evidence, the con-
clusion warranted is that survey responses are legitimate, non-
capricious expressions of opinion and that they can be relied upon
for the purposes of this research.

TABLE 10
CONTINGENCY TABLE FORMAT FOR TESTING EFFECT
OF PARTICIPATION ON EXPRESSION
OF OPINION

	Quality	
	Opinion	No Opinion
Participation		
Don't Participate		
Participate		

Comparison of survey socioeconomic characteristics
    with those for the Gainesville community

The distribution of the socioeconomic characteristics of
survey respondents was compared to the distribution of such character-
istics in the Gainesville community in order to identify the existence
of significant response biases. The x^2 statistic was used for this
purpose. Table 11 reflects the actual and expected response fre-
quencies and the related test statistics.

The expected frequencies reported in Table 11 are based on
the Census of Population and Housing--Gainesville, Florida, Standard
Metropolitan Statistical Area, 1970. In using the 1970 census, it
is assumed that the distribution of socioeconomic characteristics
in the population has not changed significantly since 1970. Further-
more in making comparisons of census and survey data, certain inter-
polations were necessary because census categories were sometimes
slightly different from those in the survey. Despite these limita-
tions, the census data is believed to be adequate for the purpose
of this research.

The Gainesville Community was operationally defined as all
census tracts within the city limits plus all census tracts contiguous
to the city limits (census tracts 1 through 17).

The hypothesis of no significant difference between survey
socioeconomic characteristics and those for the community was re-
jected for all 8 socioeconomic characteristics. The major survey
biases are listed below:

TABLE 11

COMPARISON OF SURVEY SOCIOECONOMIC CHARACTERISTICS
WITH THOSE FOR GAINESVILLE COMMUNITY

Socioeconomic Characteristics	Act. Freq.	%	Expt. Freq.	%	Calculated X	Degrees of Freedom	X^2 from Table*	Sig. Level
Sex								
Male	331	68%	242	50%				
Female	153	32%	242	50%				
	484		484		65.40	1	10.83	.001
Age								
16-20	28	6%	106	22%				
21-30	257	53%	154	32%				
31-50	118	25%	130	27%				
51-65	60	12%	58	12%				
Over 65	19	4%	34	7%				
	482		482		134.10	4	18.47	.001
Level of Education								
Grammer School	7	2%	159	33%				
High School	35	7%	116	24%				
Some College	130	27%	67	14%				
College Degree	310	64%	140	29%				
	482		482		467.50	3	16.27	.001
Married								
Yes	293	61%	231	48%				
No	187	39%	251	52%				
	482		482		32.96	1	10.83	.001

TABLE 11--Continued

Socioeconomic Characteristics	Act. Freq.	%	Act. Freq.	%	Calculated X^2	Degrees of Freedom	X^2 from Table*	Sig. Level
Number Living in Home								
1	80	17%	77	16%				
2	180	37%	154	32%				
3-5	205	43%	216	45%				
Over 5	15	3%	33	7%				
	480		480		14.89	3	11.34	.01
Family Income								
0-5,000	101	21%	124	26%				
5,001-10,000	92	19%	153	32%				
10,001-15,000	78	16%	105	22%				
Over 15,000	206	44%	95	20%				
	477		477		165.00	3	16.27	.001
College Student								
Yes	205	43%	142	30%				
No	268	57%	331	70%				
	473		473		39.94	1	10.83	.001
Live in Gainesville City Limits								
Inside City	326	68%	392	82%				
Outside City	152	32%	86	18%				
	478		478		61.76	1	10.83	.001

*Blalock, 1972, p. 569.

1. Males outnumber females 2:1 in the survey, but for the community the ratio is unity.

2. The 21-30 age group is considerably overrepresented while the 16-20 age group is considerably underrepresented.

3. While 64% of survey respondents have a college degree, only 29% of those in the community do.

4. Married people are overrepresented.

5. While 44% of the survey respondents reported family incomes over $15,000, only 20% of family incomes in the community exceed this amount.

6. College students are overrepresented.

7. Those outside the city limits are overrepresented. Socioeconomic characteristics of survey respondents are not representative of those in the community. The survey is especially biased towards those with a high level of education and income and those in the 21-30 age bracket. Community members with little formal education have provided almost no input.

While the biases found in the socioeconomic characteristics do not necessarily entail biases in ratings of importance, quality and adequacy, the burden of proving they do not must be borne by anyone who wishes to generalize from the survey to the community at large.

Effect of geographic location on ratings

In an effort to determine if recreation values, needs and experiences were uniform throughout the community, ratings of

TABLE 12
CONTINGENCY TABLE FORMATS FOR TESTING EFFECT
OF GEOGRAPHIC LOCATION ON IMPORTANCE,
QUALITY AND ADEQUACY

	Importance			Degrees of
	Low	Average	High	Freedom
Predominantly Southeast with some Northeast				
Northeast				
Predominantly Southwest				
Northwest				
Predominantly Southwest with some Northwest				
Northwest				10

	Quality			
	Poor	Fair	Good	
Predominantly Southeast with some Northeast				
Northeast				
Predominantly Southwest				
Northwest				
Predominantly Southwest with some Northwest				
Northwest				10

	Adequacy		
	Adequate	Inadequate	
Predominantly Southeast with some Northeast			
Northeast			
Predominantly Southwest			
Northwest			
Predominantly Southwest with some Northwest			
Northwest			5

TABLE 12
CONTINGENCY TABLE FORMATS FOR TESTING EFFECT
OF GEOGRAPHIC LOCATION ON IMPORTANCE,
QUALITY AND ADEQUACY

| | Importance | | | Degrees of |
	Low	Average	High	Freedom
Predominantly Southeast with some Northeast				
Northeast				
Predominantly Southwest				
Northwest				
Predominantly Southwest with some Northwest				
Northwest				10

| | Quality | | | |
	Poor	Fair	Good	
Predominantly Southeast with some Northeast				
Northeast				
Predominantly Southwest				
Northwest				
Predominantly Southwest with some Northwest				
Northwest				10

| | Adequacy | | |
	Adequate	Inadequate	
Predominantly Southeast with some Northeast			
Northeast			
Predominantly Southwest			
Northwest			
Predominantly Southwest with some Northwest			
Northwest			5

importance, quality and adequacy were compared by geographic location. X^2 was used to test the hypothesis that location has no effect. The contingency table formats used in the tests are presented in Table 12 and the results of the tests are reported in Table 13.

Geographic location was found to have no effect on importance ratings and little effect on quality. For adequacy, however, 3 of the 9 facility ratings differed by location.

Based on the above results, the separate reporting of ratings of importance and quality by geographic location would provide only minimal additional information. Furthermore, the observed bias towards certain locations (see p. 82) does not appear to be significant insofar as the measures of importance and quality are concerned. The observed differences in facility adequacy by geographic location should be reported to decision makers as such information should be useful in deciding on locations for new facilities.

TABLE 13
EFFECT OF GEOGRAPHIC LOCATION ON IMPORTANCE,
QUALITY AND ADEQUACY

	Sig.	Not Sig.
Importance	0	28
Quality	3	25
Adequacy	3	6

Note: X^2 was used to identify differences significant at .05 for 28 programs and 9 facilities.

Effect of socioeconomic characteristics on ratings

The contingency table formats used to compare importance, quality and adequacy ratings by socioeconomic characteristics are presented in Table 14. The results of the X^2 tests of the hypothesis that socioeconomic characteristics have no effect are presented in Table 15 and are summarized below:

1. Importance ratings are substantially affected by sex, age, level of education, marital status and college status. Family income and residence had some effect. Number in home had no effect.

2. Quality ratings were not substantially affected by any socioeconomic characteristic.

3. Sex, age, level of education and family income had some effect on adequacy ratings.

The biases toward certain socioeconomic characteristics (see p. 98) coupled with the effect of such characteristics on importance and adequacy ratings has several implications for this research:

1. Generalization of these ratings to the community at large is unwarranted.

2. Conclusions based on these ratings can only be tentative-- more representative ratings may yield different results necessitating different conclusions.

Because quality was not greatly affected by socioeconomic character- istics, observed biases would not appear to be as critical for quality ratings.

TABLE 14

CONTINGENCY TABLE FORMATS FOR TESTING EFFECTS
OF SOCIOECONOMIC CHARACTERISTICS ON IMPORTANCE,
QUALITY AND ADEQUACY

| | Importance | | | Degrees of |
	Low	Average	High	Freedom
Sex				
Male				
Female				2
Age				
16-20				
21-30				
31-50				
51-65				
Over 65				8
Level of Education				
No College				
Some College				
College Degree				4
Married				
Yes				
No				2
Number Living in Home				
1				
2				
Over 2				4
Family Income				
0-5,000				
5,001-10,000				
10,001-15,000				
15,001-20,000				
20,001-30,000				
Over 30,000				10

TABLE 14--Continued

| | Importance | | | Degrees of |
	Low	Average	High	Freedom
College Student				
Yes				
No				2
Live in Gainesville City Limits				
Inside City				
Outside City				2

| | Quality | | | |
	Poor	Fair	Good	
Sex				
Male				
Female				2
Age				
16-20				
21-30				
31-50				
51-65				
Over 65				8
Level of Education				
No College				
Some College				
College Degree				4
Married				
Yes				
No				2

TABLE 14--Continued

	Quality			Degrees of
	Poor	Fair	Good	Freedom
Number Living in Home				
1				
2				
Over 2				4
Family Income				
0-5,000				
5,001-10,000				
10,001-15,000				
15,001-20,000				
20,001-30,000				
Over 30,000				10
College Student				
Yes				
No				2
Live in Gainesville City Limits				
Inside City				
Outside City				2

	Adequacy		Degrees of
	Adequate	Inadequate	Freedom
Sex			
Male			
Female			1

TABLE 14--Continued

	Adequacy		Degrees of Freedom
	Adequate	Inadequate	
Age			
16-20			
21-30			
31-50			
51-65			
Over 65	_____		4
Level of Education			
No College			
Some College			
College Degree	_____		2
Married			
Yes			
No	_____		1
Number Living in Home			
1			
2			
Over 2	_____		2
Family Income			
0-5,000			
5,0001-10,000			
10,001-15,000			
15,001-20,000			
20,001-30,000			
Over 30,000	_____		5
College Student			
Yes			
No	_____		1
Live in Gainesville City Limits			
Inside City			
Outside City	_____		1

TABLE 15

EFFECTS OF SOCIOECONOMIC CHARACTERISTICS ON
IMPORTANCE, QUALITY AND ADEQUACY

Socioeconomic Characteristics	Importance		Quality		Facility Adequacy	
	Sig.	Not Sig.	Sig.	Not Sig.	Sig.	Not Sig.
Sex	16	12	0	28	2	7
Age	5	23	0	28	2	7
Level of Education	4	24	2	26	2	7
Married	9	19	3	25	0	9
Number Living in Home	0	28	1	27	1	9
Family Income	3	25	1	27	0	8
College Student	4	24	0	28	0	9
Live in Gainesville City Limits	2	26	3	25	0	9

Note: X^2 was used to identify differences significant at .05 for 28 programs and 9 facilities.

Effect of participation on ratings of importance and quality

The survey questionnaire was designed so that the opinions of program participants could be compared with non-participants in order to learn if significant differences of opinion existed between these 2 groups. The contingency tables used for the X^2 test are presented in Table 16 below. Significant differences between the ratings of participants and non-participants were found for 18 of 28 programs in regard to importance and for 8 of 28 programs in regard to quality.

TABLE 16
CONTINGENCY TABLE FORMATS FOR TESTING EFFECT OF
PARTICIPATION ON RATINGS OF IMPORTANCE AND QUALITY

Participation	Importance		
	Low	Average	High
Participate			
Don't Participate			
	Quality		
	Poor	Fair	Good
Participate			
Don't Participate			

The existence of substantial difference of opinions between participants and non-participants indicates the need for decision makers to consider both groups in making decisions regarding recreation resource allocations. Since participants have the most to

gain, they are likely to be more vocal than non-participants. If
their views are the only input to the decision making process, non-
optimal decisions may result.

Objective Measures of Input and Output Quantity

Each division head was asked to provide the researcher with
estimates of output quantity (see p. 50) and inputs (see p. 52) for
the programs in his division. In order to promote the comparability
and accuracy of these estimates, data collection forms (see Appendix D)
were prepared by the researcher for the use of the division heads.
After the forms had been completed, they were reviewed by the researcher.
Questions about the data reported on the forms were discussed with
appropriate division head. These discussions resulted in some revi-
sions of the original estimates.

The division heads were able to supply the researcher with
estimates of the data requested for 45 of the original 55 programs.
Difficulties were experienced in the development of other direct
costs (which as reported herein consist primarily of materials and
supplies):

1. Materials used by the Aquatics Division consist mainly of
chemicals for water treatment—any allocation to individual programs
would have been arbitrary.

2. For Center Division programs requiring payment of a
materials usage fee, no estimate of materials costs was provided.

3. All division heads stated that reliable estimates of
utility and maintenance costs could not be made.

The data obtained are listed in Tables 17 and 18. Table 17 contains the programs evaluated by all 3 groups. Table 18 contains the programs which were only evaluated by the GRD and PRAB.

Delphi

Data Collection

Delphi questionnaires (Appendix E) containing all 55 recreation programs and 9 facilities were given to the 14 GRD supervisors and 11 PRAB members. The cover letters accompanying the questionnaires can be found in Appendix F. All questionnaires were returned to the researcher and the following statistics were separately computed for each group: (1) means, (2) standard deviations, (3) frequency distributions, (4) histograms, (5) medians, and (6) interquartile ranges. The importance and quality means for round 1 are given in Tables 19 and 20. The median and interquartile range (feedback) for the importance and quality of each program were placed on another questionnaire (Appendix E) which was then given to the GRD and PRAB. A cover letter (Appendix F) explaining the purpose of the feedback information accompanied the questionnaires. Again all questionnaires were returned and statistics were computed for the second round responses.

Analysis

The following 4 criteria for the existence of a group judgement were discussed in Chapter IV (pp. 72-73):

1. change and convergence with feedback

2. group reliability

TABLE 17

INPUT AND OUTPUT MEASURES--PROGRAMS EVALUATED
BY ALL THREE GROUPS

Program Name	Participant Hours	Spectator Hours	Direct Labor Cost	Other Direct Costs	Staff Hours	Volunteer Hours	Total* Input Hours	Fees
Adult basketball	6,790	4,725	$ 2,460	$-	909	-	909	$ 770
Adult ceramics	8,532	-	3,157	-	714	-	714	-
Adult flag football	5,472	4,620	1,496	-	440	-	440	320
Archery activities	3,640	-	-	-	-	-	-	-
Drama workshop & play	680	75	405	5	185	-	185	-
Modern dance lessons	6,750	-	-	-	-	450	450	-
Pre-school training	448	-	474	48	112	-	112	48
Public swimming	108,000	-	28,196	-	10,560	-	10,560	17,100
Recreation center dances	11,955	-	1,307	-	319	446	765	-
Sewing lessons	4,968	-	551	110	192	408	600	-
Skin & scuba diving lessons	960	-	270	-	120	240	360	320
Springboard diving lessons	1,200	-	288	-	120	-	120	520
Square dance lessons	3,696	-	108	650	21	780	801	768
Summer track	1,260	-	330	-	120	-	120	-
Track & field day	8,800	800	-	-	160	80	240	-
Tumbling lessons	1,728	-	482	-	144	-	144	648
Youth arts & crafts	3,360	-	1,717	312	336	-	336	-
Youth gymnastics	4,080	-	2,640	100	960	-	960	-
Youth swim lessons	10,800	-	2,646	-	1,080	-	1,080	2,108
Wrestling lessons	84	-	55	-	12	-	12	4,680
Youth ceramics	3,360	-	1,717	-	336	-	336	-

*This is the sum of staff hours and volunteer hours.

TABLE 18

INPUT AND OUTPUT MEASURES--PROGRAMS EVALUATED
BY ONLY TWO GROUPS

Program Name	Partic-ipant Hours	Spec-tator Hours	Direct Labor Cost	Other Direct Costs	Staff Hours	Volun-teer Hours	Total* Input Hours	Fees
Adult arts & crafts	900	-	$ 383	$-	75	-	75	$-
Adult exercise lessons	2,304	-	643	-	192	-	192	432
Adult softball	41,736	9,400	9,165	-	3,055	-	3,055	1,965
Adult swim lessons	264	-	115	-	48	-	48	143
Art display	550	625	128	123	108	60	168	-
Baton lessons	80	-	-	-	8	-	8	-
Bowling lessons	240	-	120	-	24	-	24	120
Camping skills instruction	798	-	496	-	165	-	165	84
Cheerleading lessons	1,800	675	5	108	52	130	182	-
Competitive swimming	10,800	-	2,400	377	960	-	960	750
Cooking lessons	720	-	207	-	72	-	72	-
Duplicate bridge	10,400	-	832	-	208	-	208	-
Easter egg hunt	10,000	9,000	-	94	435	-	435	-
Girls softball	879	446	116	290	50	140	190	150
Senior citizen activities	20,380	-	2,121	-	415	-	415	-
Supervised playground activities	210,900	-	35,979	4,021	17,324	-	17,324	-
Teen nutrition classes	468	-	-	-	-	52	52	-
Tennis lessons	480	-	240	-	80	-	80	250
Youth baseball	29,848	12,180	2,436	4,320	1,044	6,400	7,444	3,200
Youth basketball	5,284	4,395	365	-	102	1,260	1,362	3,200
Youth football	14,044	9,720	810	2,275	324	2,470	2,794	1,300
Water ballet	578	-	126	-	53	-	53	66
Water safety & lifesaving instruction	960	-	90	-	36	30	66	60
Women's volleyball	900	750	140	-	68	-	68	120

*This is the sum of staff hours and volunteer hours.

TABLE 19

MEAN RATINGS OF PROGRAM IMPORTANCE AND QUALITY—EVALUATED BY
ALL THREE GROUPS

Program Name	GRD*		PRAB*		Community	
	Importance	Quality	Importance	Quality	Importance	Quality
Adult basketball[ab]	3.929	4.583	3.545	4.222	2.761	3.500
Adult ceramics	3.500	4.231	3.273	4.250	3.033	3.719
Adult flag football	4.071	4.615	3.000	4.500	2.531	3.271
Archery activities	2.846	3.600	2.818	4.000	2.919	3.413
Drama workshop & play	2.929	3.500	3.000	2.200	3.054	3.373
Golf lessons	3.231	3.500	2.636	1.000	3.104	3.509
Modern dance lessons	3.077	3.714	2.818	3.000	2.547	3.375
Park & picnic facilities[ab]	4.286	4.571	4.273	4.455	4.290	3.834
Pre-school training[ab]	3.429	3.143	2.182	4.000	3.730	3.494
Public swimming[ab]	4.143	4.429	4.545	4.545	4.184	3.642
Racquetball facilities[b]	3.786	3.769	3.364	3.333	3.85	3.476
Recreation center dances[b]	3.143	3.083	2.909	3.750	2.971	3.172
Sewing lessons	3.500	3.857	3.455	3.833	3.113	3.447
Skin & scuba diving lessons	3.214	3.833	2.818	3.333	2.579	3.542
Square dance lessons	3.571	4.500	3.000	1.667	2.508	3.488
Springboard diving lessons	3.357	3.625	2.818	3.750	2.857	3.374
Summer track[ab]	3.786	3.636	3.818	3.889	3.099	3.581
Track & field day[ab]	3.714	4.462	3.727	4.333	3.408	3.746
Tumbling lessons[ab]	3.429	3.846	3.091	3.714	2.832	3.287
Youth arts & crafts	3.571	3.769	3.636	4.400	3.829	3.900
Youth ceramics	3.357	3.636	3.182	3.500	3.330	3.821
Youth gymnastics[ab]	3.429	3.538	3.455	3.667	3.668	3.902
Youth swim lessons[ab]	4.286	4.643	4.364	4.455	4.238	3.899
Wrestling lessons	2.929	2.800	2.727	2.500	2.713	2.894

*Based on first round responses.

[a]Included in revised three groups multivariable-multirater matrix.

[b]Included in revised two group multivariable-multirater matrix.

TABLE 20

MEANS RATINGS OF PROGRAM IMPORTANCE AND QUALITY--EVALUATED
BY ONLY TWO GROUPS

Program Name	GRD*		PRAB*	
	Importance	Quality	Importance	Quality
Adult arts & crafts[b]	3,214	3,833	3,273	4.167
Adult exercise lessons	3,429	3,778	3,545	3.800
Adult gymnastics[b]	2,857	3,000	2,455	2.750
Adult softball[b]	4,500	4,692	3,818	4.889
Adult swim lessons	4,000	4,429	3,727	3.833
Art display	3,429	4,100	3,455	4.000
Baton lessons	3,286	4,000	2,727	3.833
Bowling lessons	2,929	3,375	3,091	2.833
Camping skills instruction	3,000	3,000	3,182	2.250
Cheerleading lessons[b]	3,071	3,769	2,727	4.333
Competitive swimming	3,643	4,429	3,545	4.300
Cooking lessons	2,857	4,000	2,818	3.000
Duplicate bridge	3,214	3,800	2,900	3.250
Easter egg hunt	2,857	3,462	3,545	4.333
Girls softball[b]	4,071	4,000	3,727	4.000
Racquetball tournaments	3,214	3,444	2,909	3.000
Recreation center facilities	4,000	4,000	4,182	3.900
Recreation center games[b]	3,714	3,615	3,727	3.875
Senior citizen activities	4,538	4,167	4,091	4.625
Supervised playground activities[b]	4,071	4,667	4,182	4.100
Swim meets[b]	3,357	4,571	3,818	4.727
Teen nutrition classes	2,786	3,600	2,091	1.000
Tennis facilities[b]	4,429	4,714	4,091	4.545
Tennis lessons	4,077	4,182	4,091	3.889
Tennis tournaments[b]	3,429	4,385	3,909	3.455
Youth baseball[b]	4,643	4,833	4,273	4.455
Youth basketball[b]	4,214	4,462	4,000	4.000

TABLE 20—Continued

Program Name	GRD*		PRAB*	
	Importance	Quality	Importance	Quality
Youth football[b]	4.357	4.615	4.333	4.444
Water ballet	2.714	3.375	2.300	2.000
Water safety & lifesaving instruction[b]	3.929	4.643	4.455	4.500
Women's volleyball	3.500	3.900	3.545	2.333

*Based on first round responses.

[b]Included in revised two group multivariable-multirater matrix.

3. reasonable distribution

4. consensus in group

All 4 criteria were applied to the opinions of the GRD and PRAB and the last 3 were applied to the opinions of the community.*
The results and conclusions follow.

Change and convergence on iteration with feedback

This criterion requires shifts of individual responses toward the group response and reduction in group variability. Based on the results of previous research with group value judgements (Dalkey et.al., 1972),

1. significant changes in mean ratings were not expected between rounds

2. high correlations between round 1 and 2 means were expected

3. standard deviations were expected to decrease from round 1 to round 2.

The changes which occurred in program importance and quality means and standard deviations between round 1 and 2 are presented in Table 21.

While changes in means occurred for about 85% of the programs, a t test of the statistical significance of the changes revealed that none of them were significant. Pearson product-moment correlation coefficients between the 55 program means of round 1 and 2 were high:

---

*Since the opinions of community members were obtained but once, criterion 1 was not applicable.

TABLE 21

NATURE OF CHANGE IN IMPORTANCE AND QUALITY MEANS AND
STANDARD DEVIATIONS FROM ROUND ONE TO ROUND TWO

| | Gainesville Recreation Department | | | | Public Recreation Advisory Board | | | |
| | Means | | Standard Deviations | | Means | | Standard Deviations | |
	Importance	Quality	Importance	Quality	Importance	Quality	Importance	Quality
Decrease	13	14	35	36	23	27	29	36
Increase	31	35	16	13	22	21	18	12
No Change	11	5	4	5	10	6	8	6
	55	54	55	54	55	54	55	54

	GRD	PRAB
Importance	.9777*	.9575*
Quality	.9422*	.9533*

For about one-third of the programs, standard deviations failed to decrease between rounds. In view of this, an overall assessment of convergence was made by comparing the mean standard deviations of program importance and quality between rounds:

	GRD		PRAB	
	Rd. 1	Rd. 2	Rd. 1	Rd. 2
Importance	.7356	.6712	.8853	.8265
Quality	.7548	.6621	.9070	.7398

Since round 1 standard deviations exceed round 2, some convergence has occurred. Except for the importance standard deviations of the PRAB, the differences are statistically significant ($\underline{t}$ test).

In summary, mean ratings behaved as anticipated and convergence, while less than expected, did occur. The criterion of "change and convergence on iteration with feedback" has been met fairly well.

Since no significant changes in mean ratings occurred between rounds, additional iterations did not appear fruitful. Furthermore, since the information content of round 1 and 2 ratings are essentially the same, only round 1 ratings will continue to be used.

---

*Significant at .001.

Group reliability

Dalkey et al. state that

> Given two similar groups (e.g., two groups selected out of
> a larger group at random) the group judgements on a given
> question should be similar. Over a set of such value judge-
> ments, the correlation for the two subgroups should be high.
> (1972, p. 57)

For this research the similar groups would appear to be the GRD and
PRAB and community groups A and B. The Pearson product-moment cor-
relation coefficients between ratings for these 2 sets of groups are
given below:

	GRD & PRAB (55 programs)	Community A & B (4 programs)
Importance	.7961 (.001)*	.9197 (.04)*
Quality	.6345 (.001)*	.8451 (.08)*

Based on the above, the reliability criterion would appear to be met.

Reasonable distribution

Concerning this criterion Dalkey et al. write

> If the distribution of group response on a given numerical
> value judgement is flat, indicating group indifference, or
> if it is U-shaped, indicating either that the question is
> being interpreted differently by two subgroups, or there is
> an actual difference of assessment by two subgroups, then
> it seems inappropriate to assert that the group considered
> as a unit has a judgement on that question.** (1972, p. 57)

---

*Significance level

**Based on prior Delphi research, a single peaked, normal type dis-
tribution is expected if a group judgement exists.

The frequency distributions (histograms) for program importance and quality for the GRD, PRAB and community were examined and judged to be flat, bimodal or single peaked.* The results are reflected in Table 22. For importance 96% of the programs satisfy the criterion of "reasonable distribution" while for quality 80% of the programs satisfy it.

In summary, the criterion of "reasonable distribution" has been met for most programs.

Consensus in group

In the development of the social service measurement model for the JCF (see p. 30) consensus among group members was used as a criterion for group judgement. Unless approximately 80% of the group agreed on 2 contiguous categories of the rating scale, a group judgement was not considered to exist.

The highest percentage of group members expressing an opinion in 2 contiguous categories for program importance and quality ratings are presented in Table 23. The GRD exhibits the greatest degree of consensus and is followed by the PRAB. While consensus is lowest for the community, in no case is it less than 50%.

The degree of consensus for the GRD and PRAB at the end of round 2 compares very favorably with the degree of consensus achieved

---

*Based on prior Delphi research, a single peaked, normal type distribution is expected if a group judgement exists.

TABLE 22

TYPE OF DISTRIBUTION FOR
PROGRAM IMPORTANCE AND QUALITY RATINGS

Type of Distribution	GRD*		PRAB*		COMMUNITY	
	Importance	Quality	Importance	Quality	Importance	Quality
Flat	0	5	1	12	0	1
Bimodal	1	3	5	5	0	1
Single Peaked	54	47	49	33	28	26
	55	55	55	55	28	28

* Round 1.

TABLE 23

DEGREE OF GROUP CONSENSUS FOR PROGRAM
IMPORTANCE AND QUALITY RATINGS

| | Gainesville Recreation Dept. | | | | Public Recreation Advisory Board | | | | Community | |
| | Impt. | | Qual. | | Impt. | | Qual. | | Impt. | Qual. |
Highest % of Agreement in Two Contiguous Categories	Rd. 1	Rd. 2	Rd. 1	Rd. 2	Rd. 1	Rd. 2	Rd. 1	Rd. 2		
90-100%	18	27	25	31	13	18	25	25	0	0
80-89%	15	17	13	14	15	15	12	18	2	0
70-79%	20	10	10	6	18	16	10	4	9	3
60-69%	1	1	5	3	7	5	9	5	17	19
50-59%	1	-	1	-	1	1	5	2	-	6
Below 50%	-	-	-	-	1	-	-	-	-	-
	55	55	54	54	55	55	54	54	28	28

for the JCF at the end of round 4* (Reisman et al., 1970, pp. 21-23)!

In summary the criterion of consensus appears to have been met reasonably well.

Overall conclusion

It is the conclusion of this researcher that a group judgement exists for most of the programs being evaluated.

Use of Multivariable-Multirater Matrix to Assess the Validity of Measures of Importance and Quality

In this section the results of using a multivariable-multirater** (M-M) methodology to evaluate the validity of the measures of program importance and quality are discussed. The following information is applicable to all 4 of the M-M matrices to be discussed.

Pearson's product-moment correlation was used to produce the coefficients for the matrices. Reliabilities are in parenthesis and convergent validity coefficients (hereafter validities) are underlined. Each monorater triangle is enclosed by broken lines and each heterorater block is enclosed by solid lines. In referring to heterovariable correlations, the column name will always precede the row name.

---

*The reason for only 2 rounds in this research was stated at page 120.

**Based on the multitrait-multimethod methodology of Campbell and Fiske (1959).

The evaluations are based on criteria set forth by Campbell and Fiske (1959). These criteria were discussed in Chapter IV and are restated below specifically for this research:

A. Reliabilities should be high and should exceed all other correlation coefficients.

B. Convergent validity is demonstrated by high validities.

C. Evidence for discriminant validity is provided by 3 criteria:

1. A validity coefficient should be higher than the coefficients lying in its column and row of the heterorater block.

2. A validity coefficient should be higher than the quality-value coefficients in the monorater triangles.

3. The same pattern of variable interrelationships should be shown in all of the heterorater blocks.

In the evaluation of the individual matrices, reliability, convergent validity and discriminant validity will be considered. For discriminant validity, the 3 separate requirements will be designated by the numbers 1, 2, and 3 respectively.

Three Group Matrix

The M-M matrix in Table 24 contains the correlation coefficients for the 24 programs (see Table 19, p. 115) evaluated by the GRD, PRAB and community.

Reliability. Reliabilities for the GRD and PRAB are based on

TABLE 24

THREE GROUP MULTIVARIABLE-MULTIRATER MATRIX

		Opinions of Recreation Supervisors		Opinions of Advisory Board Members		Opinions of Community	
		Program Impt.	Program Quality	Program Impt.	Program Quality	Program Impt.	Program Quality
Opinions of Recreation Supervisors	Program Impt.	(.9719*)	.7837*	.7343*	.6260*	.5407*	.5051*
	Program Quality		(.9494*)	.5781*	.5609*	.1803**	.4753*
Opinions of Advisory Board Members	Program Impt.			(.9575*)	.5402*	.6558*	.5573*
	Program Quality				(.9639*)	.3311**	.3932*
Opinions of Program Users	Program Impt.					(.9197*)	.6261*
	Program Quality						(.8451**)

Notes: Except for reliabilities of community, n = 24. For community reliabilities, n = 4.

*Significant at .05.

**Not significant at .05.

correlations between round 1 and 2 program means.* They are quite
high and suggest that the measures of importance and quality are
stable. Since community members only evaluated the programs 1 time,
traditional reliabilities could not be computed. However, some indi-
cation of the stability of the community measures could be obtained
from the correlations between the means of the programs commonly
evaluated by the A and B groups. These correlations are high, again
suggesting stability of the measures. The reliabilities exceed
all other coefficients in the matrix.

Convergent validity. All validities, except quality (.3932) for
PRAB and community, are of sufficient size and significance as to
provide evidence of convergent validity. Importance validities exceed
those for quality and the highest validities occur between the GRD
and PRAB.

Discriminant validity.

1. While importance validities exceed the coefficients in the
column and row of the heterorater blocks, the validities for quality
do not.

2. Validities do not exceed all the heterovariable coefficients
in the monorater triangles.

3. In all heterorater blocks, the importance validities are
the highest and the quality-value coefficients are the second highest.
No other consistent interrelationships exist.

---

*Subsequent reliabilities for GRD and PRAB were computed in a similar
manner.

Summary. Evidence for convergent validity exists. While there is some evidence of discriminant validity for importance, there is none for quality. Importance and quality are highly correlated.

Two Group Matrix

The M-M matrix in Table 25 contains the correlation coefficients for the 55 programs (see Tables 19 and 20, p. 115) evaluated by the GRD and PRAB.

Reliability. Reliabilities are high and exceed all other coefficients in the matrix.

Convergent validity. All validities are high and provide evidence of convergent validity. The importance validity coefficient exceeds the 1 for quality.

Discriminant validity.

1. The importance validity coefficient exceeds the coefficients in its row and column of the heterorater block, but the quality validity coefficient does not.

2. The importance validity coefficient exceeds the heterovariable coefficient in the monorater triangles. The quality validity coefficient does not.

3. Not applicable.

Summary. Evidence for both convergent and discriminant validity exists for importance. Evidence for convergent validity but not discriminant validity exists for quality. Importance and quality are highly correlated.

TABLE 25

TWO GROUP MULTIVARIABLE-MULTIRATER MATRIX

	Opinions of Recreation Supervisors		Opinions of Advisory Board Members	
	Program Impt.	Program Quality	Program Impt.	Program Quality
Opinions of Recreation Supervisors				
Program Impt.	(.9777*)	.7706*	.7961*	.6339*
Program Quality		(.9422*)	.6789*	.6345*
Opinions of Advisory Board Members				
Program Impt.			(.9645*)	.6521*
Program Quality				(.9533*)

Note: In all cases n = 55.
 *Significant at .05.

Revised Multivariable-Multirater Matrices

The correlations in the preceding matrices were based on all
programs evaluated. However, as noted previously (pp.121 -125), certain
programs failed to meet the Delphi criteria for the existence of a
group judgement. In addition to the Delphi criteria, the require-
ment that at least a majority of GRD and PRAB members express an
opinion on a program was also imposed. This criterion was based on
the belief that a group judgement could not be said to exist unless
a majority of group members expressed an opinion. A question of
interest is would convergent and discriminant validity improve if
those programs not satisfying the above criteria were eliminated?

Using the above criteria, revised 3 and 2 group sets of pro-
grams were prepared. (Those programs designated by "a" in Tables
19 comprise the 3 group set; those programs designated by "b" in
Tables 19 and 20 comprise the 2 group set). The M-M matrices for
the revised programs are shown in Tables 26 and 27 respectively.

Three Group Revised Matrix

Reliability. Reliabilities are high for GRD and PRAB and
exceed all other coefficients. Reliabilities for community were
not calculated as 2 of the 4 commonly evaluated programs were elim-
inated.

Convergent validity. All validities, except quality for
GRD and community, are high and provide evidence of convergent
validity. Importance validities exceed those for quality.

TABLE 26

THREE GROUP REVISED MULTIVARIABLE-MULTIRATER MATRIX

		Opinions of Recreation Supervisors		Opinions of Advisory Board Members		Opinions of Community	
		Program Impt.	Program Quality	Program Impt.	Program Quality	Program Impt.	Program Quality
Opinions of Recreation Supervisors	Program Impt.	(.9523*)	.6929*	.8993*	.8473*	.8394*	.8041*
	Program Quality		(.9728*)	.5152**	.8659*	.5491***	.6331**
Opinions of Advisory Board Members	Program Impt.			(.9208*)	.8467*	.9167*	.7970*
	Program Quality				(.9496*)	.8066*	.8009*
Opinions of Program Users	Program Impt.					***	.8153*
	Program Quality						***

Notes: In all cases, n = 7.

*Significant at .05; **not significant at .05.

***Unable to calculate.

Discriminant validity.

1. Importance validites exceed the coefficients in the columns and rows of the heterorater blocks. The quality validity coefficient for the GRD and PRAB exceeds the coefficients in its column and row of the heterorater block. The remaining quality validities do not.

2. The validities for the GRD and PRAB and the importance validity coefficient for the PRAB and community exceed the quality-importance coefficients in the monorater triangles. The other validities do not.

3. Importance validities exceed all other coefficients in the heterorater blocks. No other consistent interrelationships exist.

Summary. Evidence for convergent validity exists. Evidence of discriminant validty for importance exists. While some evidence of discriminant validity for quality exists, it is not strong. Importance and quality are highly correlated.

Two Group Revised Matrix

Reliability. Reliabilities are high and exceed all other coefficients.

Convergent validity. Validities are high and provide evidence of convergent validity. The importance validity coefficient exceeds the 1 for quality.

Discriminant validity.

1. Validities exceed coefficients in column and row of heterorater block.

TABLE 27

TWO GROUP REVISED MULTIVARIABLE-MULTIRATER MATRIX

		Opinions of Recreation Supervisors		Opinions of Advisory Board Members	
		Program Impt.	Program Quality	Program Impt.	Program Quality
Opinions of Recreation Supervisors	Program Impt.	(.9693*)	.6609*	.7437*	.4098*
	Program Quality		(.9538*)	.6408*	.6473*
Opinions of Advisory Board Members	Program Impt.			(.9237*)	.4399*
	Program Quality				(.9352*)

Notes: In all cases, n = 22.

*Significant at .05.

2. Importance validity coefficient exceeds quality-importance coefficients in monorater triangles. Quality validity coefficient exceeds quality-importance coefficient in monorater triangle for PRAB but not the one in monorater triangle for GRD.

3. Not applicable.

Summary. Evidence of convergent and discriminant validity exists for both importance and quality. However, correlations between importance and quality are still large.

Although elimination of programs not meeting the criteria for a group judgement had little effect on convergent validity, discriminant validity improved and quality reflects some discriminant validity. However, the correlations between importance and quality are still high. While the changes observed are not sufficient to warrant a conclusion that a significant improvement in validity has occurred, they are encouraging enough to warrant further research.

Conclusion as to Validity of Measures of Importance and Quality

Subject to the limitations imposed by response biases (see pp. 78 and 98), the following overall conclusions are derived from the results of the M-M evaluations. The importance measures possess both convergent validity and discriminant validity. Convergent validity exists for the quality measures, but it is not as strong as it is for importance. While there is some evidence of discriminant validity for quality, it is unsatisfactory. These results indicate that, as measured, importance is more valid than quality.

These findings probably reflect the more stable and uniform nature of importance which is presumed to be the result of cultural values developed over a fairly long period of time. Quality, unlike importance, is more likely affected by recent experiences which can be expected to vary from individual to individual. One implication of this is that a better measure of quality could be obtained if its measurement occurred at the end of each program (analogous to teacher-course evaluations at the end of each quarter).

Facility Adequacy

The mean ratings of facility adequacy for the GRD, PRAB and the community are presented in Table 28. Pearson's product—moment correlation coefficient was utilized to assess the agreement between the 3 groups. The coefficients are presented below:

	PRAB	Community
GRD	-.20	.42
PRAB		-.13

None of the correlation coefficients is significant at .05. There is very little agreement on the adequacy of facilities. This lack of agreement strongly implies a need for the GRD to thoroughly assess the community's recreational facility needs prior to deciding on new facilities or expansion of old ones.

The correlation between the adequacy ratings of community groups A and B is .9624 (significant at .001). This very high cor-relation between 2 independent samples of the same population provides

TABLE 28

MEAN RATINGS OF THE ADEQUACY OF FACILITIES

Facility Name	GRD	PRAB	Community A	Community B	Community A and B
Archery	1.214	1.400	1.372	1.295	1.335
Baseball	1.357	1.727	1.171	1.177	1.174
Parks	1.357	1.455	1.500	1.553	1.525
Playgrounds	1.214	1.455	1.463	1.426	1.445
Racquetball & handball	1.714	1.455	1.605	1.574	1.590
Recreation centers	1.286	1.727	1.606	1.479	1.546
Softball	1.429	1.455	1.167	1.135	1.152
Swimming pools	1.429	1.545	1.671	1.620	1.647
Tennis	1.714	1.455	1.632	1.643	1.637

evidence of the stability of the community adequacy measures and
of the comparability of the A and B groups.

The relationship between facility adequacy and quality of
programs was examined for those facilities and programs which, in
the opinion of the researcher, could be expected to be interdependent.
The facilities and programs are

Parks—park and picnic facilities

Swimming pools – skin and scuba diving lessons; springboard
diving lessons; public swimming; youth swim
lessons

Racquetball and handball—racquetball facilities

In each case of rejection the facility was almost identical to the
program. Of those for which the null hypothesis could not be re-
jected, for only 1 (archery) were the facility and program identical.
For the other 4, the facility is merely used in providing the program.
While the paucity of facility-program combinations precludes general
conclusions, the test results suggest that there are different types
of programs. For programs in which the facility is all-important,
the adequacy of that facility is a significant factor in the deter-
mination of quality. For programs in which other factors are involved
and may predominate (for example the instructor or the band), facility
adequacy is not a significant factor in the determination of the
quality ratings.

Validity of Reliance on Self-Evaluations

Since program evaluation in recreation has been limited primarily to self-evaluation studies, 1 purpose of this research was to assess the validity of exclusive reliance on information from self-evaluations. The assessment was made by comparing the measures of program importance and quality and facility adequacy with those produced by the PRAB and the community.

In previous sections, the measures of individual program importance and quality (p. 125) and facility adequacy (p. 136) produced by the GRD were correlated with those produced by the PRAB and the community. While for importance and quality significant correlations were observed, the amount of unexplained variance is still substantial as Table 29 indicates.

TABLE 29
VARIANCE IN GRD MEASURES--
EXPLAINED AND UNEXPLAINED
BY PRAB AND COMMUNITY
MEASURES

		Variance	
		Explained	Unexplained
	r	r^2	$1-r^2$
GRD & PRAB			
(55 programs)			
Importance	.79	.63	.37
Quality	.63	.40	.60
GRD & Community			
(24 programs)			
Importance	.54	.29	.71
quality	.48	.22	.78

For the measures of facility adequacy, no significant correlations were observed. These findings suggest that the information content of measures produced by the PRAB and community may be fairly high.

A comparison of average GRD ratings with those of the PRAB and community was also made. Because of the GRD's closeness to recreation activities and vested interest in its performance, it was hypothesized that the measures of importance and quality produced by the GRD would exceed those of the other groups. In order to test this hypothesis, a t test was used to compare the average of the program importance means and the average of the program quality means for the GRD with those for the PRAB and the community. The average of the program means for the GRD and the community (based on 24 programs) and for the GRD and the PRAB (based on 55 programs) are presented in Table 30. The symbols in parenthesis below the averages were used in the following statement of the null and alternative hypotheses:

Null hypotheses: $U_{1I} = U_{2I}$

$U_{1I} = U_{3I}$

$U_{1Q} = U_{2Q}$

$U_{1Q} = U_{3Q}$

Alternative hypotheses: $U_{1I} > U_{2I}$

$U_{1I} > U_{3I}$

$U_{1Q} > U_{2Q}$

$U_{1Q} > U_{3Q}$

TABLE 30
AVERAGE RATINGS OF PROGRAM IMPORTANCE AND QUALITY

	GRD	Community
Importance	3.5214 (U_{1I})	3.2224 (U_{3I})
Quality	3.8701 (U_{1Q})	3.5275 (U_{3Q})
	GRD	PRAB
Importance	3.5608 (U_{1I})	3.3997 (U_{2I})
Quality	3.9586 (U_{1Q})	3.7025 (U_{2Q})

The results of the $\underline{t}$ tests were:

Reject: $U_{1I} = U_{3I}$

$U_{1Q} = U_{2Q}$

$U_{1Q} = U_{3Q}$

Accept: $U_{1I} = U_{2I}$

For all 4 comparisons, direction is consistent with the alternative hypothesis. With 1 exception the $\underline{t}$ tests provide support for the existence of favorable bias.

The results of both the individual and overall comparisons indicate that information from self-evaluation studies should be supplemented with information from sources independent of those being evaluated. Ideally the incremental value of this information would be determined and equated with its incremental cost.

Input-Output Relationships

As discussed previously (pp. 53 and 75), linear regression analysis was to be used in an attempt to describe the nature of the relationships between labor inputs and the quantity and quality of recreation output. The simple linear model applied to the data is illustrated by the following equation:

$$y = a + bx + e$$

where y = dependent variable

x = independent variable

a = y intercept

b = change in y with respect to a change in x (slope of regression line)

e = error term (difference between the predicted y ($\hat{y}$) and the observed y); since it is assumed that the expected value of e is 0, e will drop out of the actual prediction equation.

The equations for which a and b will be estimated are

E1. $U = a + bL$; n = 45

E2.

E2.1 $U_1 = a_1 + b_1L_1$; Q = low quality; n = 6

E2.2 $U_2 = A_2 + b_2L_2$; Q = average quality; n = 25

E2.3 $U_3 = A_3 + b_3L_3$; Q = high quality; n = 13

E3. $Q = a + b\left(\frac{L}{U}\right)$; n = 45

where U = participant hours (output quantity)

L = labor hours (input)

Q = program quality

L/U = ratio of labor hours to participant hours

n = number of programs for which preceding measures
were obtained

The data for participant hours, labor hours and quality are given

in Tables 17 and 18 (p. 113). The quality measures used are those

of the GRD.* The L/U ratio was computed by dividing labor hours by

participant hours. For the equations in E2, the observed range of

program quality was divided into 3 categories: low quality (2.8 - 3.4),

average quality (3.5 - 4.2) and high quality (4.3 - 4.8). The par-

ticipant hours and labor hours associated with programs in each quality

category provide the data for equations E2.1, E2.2 and E2.3.

For equations E1 through E3, the following was obtained:

1. Scattergram in which the dependent variable is plotted

against the independent variable

2. Statistical tables containing

r and r^2
standard error of estimate
least squares estimate of "a"
least squares estimate of "b"
standard error of "b"
F value and level of significance

3. Plot of standardized residuals against predicted standard-

ized dependent variable

---

*The quality measures of the GRD were used for the following reasons:
(1) no 1 group's (GRD, PRAB and community) quality measures have been
demonstrated to be superior to the other 2 groups; and (2) since the
GRD produced the measures of labor hours and participant hours, all
measures used for the regression analyses will have the same source.

Except for the scattergrams, this information is presented in Appendix
H. The results of applying the simple linear models to the data will
now be discussed.

E1. U = a + bL

The estimated prediction equation for all 45 programs is U = -611
+ 10.7L. The r^2 of .92 is large and significant at .001. However, the
presence of several extremely large values (see Figure 4) casts doubt
on the meaningfulness of the preceding prediction equation and its
r^2.* After removing the 6 most extreme values,* the r^2 drops to .46
and the prediction equation becomes U = 1,120 + 7.03L. The 95% confi-
dence interval for U is U - 5,512<U<U + 5,512. Considering the range
in U (264 - 11,955), this confidence interval is very large and indi-
cates the impreciseness of the relationship between U and L. An
examination of the scattergram (Figure 5) reveals a weak linear re-
lationship, especially for values of L greater than 400 hours.

E2. U = a + bL

E2.1 Low quality

The scattergram for low quality (Figure 6) reveals 1 extreme
outlier. After removing it, an r^2 of .74 is obtained and the pre-
diction equation is U = 155 + 3.76L. The linear relationship (see

---

*Blalock points out that a few extreme values may produce a high r (and
therefore r^2) where none exists among the other values; he states that
where it is not feasible empirically to include more extreme values,
consideration should be given to excluding the extreme values and to
reporting the range of varability (1972, pp. 381-383).

Figure 4. Scattergram showing relationship between participant hours and labor hours for 45 programs.

210900

189818

168736

147654

126572

105490

84408

63326

42244

21162

```
                                                              *
                          *
                                                      *
            *
  *
2 *?2 *2 *
3  **
999* **
```

0   1732   3465   5197   6930   8662   10394   12126   13859   15592   17324

Figure 5. Scattergram showing relationship between participant
hours and labor hours for 39 programs.

Figure 6. Scattergram for low quality group showing relationship between participant hours and labor hours for 6 programs.

Figure 7) is weak and a 95% confidence interval for U, U − 328<U<U + 328,
is almost as large as the range of U (84 − 798).

E2.2 Average quality

The scattergram for average quality relfects 1 extreme outlier
(Figure 8). After removing it, an r^2 of .34 is obtained and the pre-
diction equation is U = 798 + 7.66L. The linear relationship is weak
(see Figure 9) and a 95% confidence interval for U, U − 5,140<U<U + 5,140,
is as large as the range of U (80 − 10,400).

E2.3 High quality

The scattergram for high quality reflects several extreme outliers
(Figure 10) and when they are removed an r^2 of .48 is obtained. The
prediction equation is U = 3,839 + 3.63L and the 95% confidence inter-
val for U is U − 6,494<U<U + 6,494. While the relationship between U
and L appears to be linear, it is very weak (Figure 11).

E3. $Q = a + b \left(\frac{L}{U}\right)$

The r^2 for the 45 programs is .001 and is totally insignificant.
The scattergram (Figure 12) shows no relationship (linear or non-linear)
between quality and the labor hour-participant hour ratio. In the
prediction equation $Q = 3.96 − .05 \left(\frac{L}{U}\right)$, the hypothesis that b = 0 cannot
be rejected (F = .002 and is not significant).

Overall Conclusion

Based on the preceding analyses, the following tentative
conclusions are presented:

Figure 7. Scattergram for low quality group showing relationship between participant hours and labor hours for 5 programs.

727

655

584

512

441

370

298

227

155

84

27   43   58   73   89   104   119   134   150   165

Figure 8. Scattergram for average quality group showing
relationship between participant hours and
labor hours for 25 programs.

Figure 9. Scattergram for average quality group showing relationship between participant hours and labor hours for 24 programs.

Figure 10. Scattergram for high quality group showing relationship
between participant hours and labor hours for 13
programs.

189836

168773

147709

126646

105582

84518

63455

42391

21328

264

1776   3503   5231   6958   8686   10414   12141   13869   15596   17324

Figure 11. Scattergram for high quality group showing relationship between participant hours and labor hours for 9 programs.

12666

11288

9910

8532

7154

5776

4398

3020

1642

264

323 597 872 1146 1421 1695 1970 2245 2519 2794

Figure 12.   Scattergram showing relationship between quality and
the labor hour-participant hour ratio.

1. While $\frac{dU}{dL} > 0$ as expected, the linear relationship* between U and L is very weak and the prediction equations are too imprecise for use in decision making. Segregating programs by level of quality (E2) did not result in a significant improvement in the strength of the relationship between U and L.

2. It appears that $\frac{dQ}{d\left(\frac{L}{U}\right)} = 0$. There is no apparent relationship, linear or non-linear, between quality and the labor hour-participant hour ratio.

3. The unsatisfactory relationships may well be due to the fact that each program is unique. Such being the case, the analyses performed are inappropriate and the relationship between inputs and outputs needs to be assessed program by program. For this type of analysis, different levels of input and output for each program are needed. Since this research was confined to 1 year, the data needed for such an analysis was not available.

---

*Based on examination of scattergrams and residual plots, non-linear relationships are not appropriate.

CHAPTER VI

CONCLUSIONS

Introduction

The overall purpose of this research was to provide evidence
as to the feasibility and efficacy of performance measurement in the
NFP area. In order to accomplish this purpose, existing methodologies
for NFP performance measurement were reviewed and assessed (Chapter
II); a performance measurement model for the GRD was developed (Chap-
ter III); and the data specified by the model was collected and analyzed
(Chapter V). Involved in the data collection and analysis were new
applications of the Delphi technique and the multitrait-multimethod
methodology and an assessment of the usefulness of self-evaluation
studies in performance measurement.

In this chapter the conclusions reached in regard to the above
will be set forth and directions for future research activity will be
discussed.

Performance Measurement Methodologies

Need for a Comprehensive Work on Performance Measurement

In this research an attempt was made to identify methodologies
which have been proposed for measuring performance in the NFP area.
The identification process revealed that while numerous methodologies
exist, they have generally been treated in isolation from one another.
A comprehensive work on NFP performance measurement is needed and

would be of considerable value to NFP administrators and researchers.
Also needed is (1) a thorough survey of both the situations in which
the methodologies have been applied and (2) the results of these ap-
plications. The survey should be designed so as to permit the deter-
mination of the state of development of performance measurement by type
(government, religious, educational, etc.) and level (e.g., govern-
ment could be classified as federal, state and local) of NFP organization.

An Improved Performance Measurement System

The review and assessment of NFP performance measurement method-
ologies resulted in the conclusion that certain methodologies are com-
plementary and can be combined into an improved performance measure-
ment system. The methodologies are (1) experimental and quasi-experi-
mental research designs (see p. 26) and cost-benefit analysis (see p.
14). The first method provides the most valid and effective way to
determine a program's impact, but it cannot be used to determine the
program's social desirability. The second method is the best way known
of assessing the social desirability of a program, but it cannot be
used to determine the program's impact. Since both impact and social
desirability should be established before implementation of a program,
the methods should be used in conjunction with one another.

Using the combined system, the impact of a program would first
be determined by use of experimental or quasi-experimental research
designs in a pilot study of the proposed program. Assuming the impact
anticipated was found to exist, a cost-benefit analysis could be used
to assess the social desirability of the program. If the program was

found to be both effective and socially desirable, it would be implemented. Its implementation should be conducted within an experimental (or quasiexperimental) framework in order to control for the effects of confounding variables. Periodically the program should be reevaluated by comparing actual with projected effects, costs and benefits. Material discrepancies should be investigated with a concomitant refinement of methodological assumptions and projections.

Feasibility of Performance Measurement

The review and assessment of the current state of performance measurement in the NFP area revealed a widespread belief that performance measurement is both feasible and efficacious. This belief is supported by the existence of several operational performance measurement methodologies (cost-benefit analysis; cost-effectiveness analysis; PPB; experimental and quasiexperimental research designs for program evaluation). While implementation of these methodologies is not yet widespread, the potential appears great. The lack of implementation appears to be due to

1. lack of administrative familiarity with methodologies

2. reluctance of administrators to be evaluated

3. lack of information systems to supply the data needed to implement the methodologies.

One approach to reducing problems 1 and 2 is for researchers interested in the NFP area to work with NFP administrators in order to inform them of the benefits and limitations of performance measurement methods and assist them in the applications of these methods.

A first step towards eliminating the third obstacle to implementation is to formally identify the activities (programs, projects, etc.) of NFP entities and to develop measures of input and output for each activity. While traditional methods (e.g., a cost accounting system) appear adequate for measuring inputs, output measurement will sometimes require the use of more novel methods (for example, Delphi and the multitrait-multimethod methodology). Given input and output measures, an information system to provide these measures to decision makers can be established.

Performance Measurement Model

Summary of Model Development

With the cooperation of the GRD, a performance measurement model was developed for the GRD. Since the model developed (Figure 2, p. 45) goes well beyond the traditional budgetary model characteristic of most NFP organizations, the ability to collect the data specified by the model was of primary concern to the researcher. The data collection focused on measures of input and output for GRD recreation programs. Since 2 of these output measures (importance and quality) were subjective in nature, their validation was a key research objective.

With the assistance of GRD supervisors, major recreation programs were identified. For most of these programs, the GRD was able to provide estimates of the following objective inputs and outputs:

1. direct costs

2. labor hours of input

3. participant and spectator hours (quantity of output)

4. user fees

The subjective output measures of program importance and quality were produced from opinions of GRD supervisors, PRAB members and a sample of residents in the Gainesville community. These same groups also provided measures of facility adequacy.

The validity of the measures of program importance and quality was assessed by means of the Delphi technique and the multitrait-multimethod methodology (multivariable-multirater). The facility adequacy measures of the 3 groups were compared for agreement. The relationships between the quantity and quality of program output and labor inputs were examined by means of regression analysis.

Conclusions

Objective input-output measures

The "program" was found to be a useful unit of account for the measurement of inputs and outputs. While a lack of research resources precluded the measurement of objective program inputs and outputs as well as desired, the active participation of GRD supervisors in providing this information indicates the potential for formally collecting it. The researcher is convinced that with sufficient resources, valid and reliable objective input-output measures can be obtained.

Subjective output measures

Based on Delphi criteria, valid group judgements of program importance and quality were found to exist. Based on the multitrait-multimethod (multivariable-multirater) criteria, the measures of program importance and quality produced from the group judgements are believed to be valid. The success achieved in measuring and validating the

importance and quality of recreation programs suggests that subjective measures are well within the purview of current methodology. To the extent that subjective measures are found useful to NFP decision makers, they can and should be provided.

Facility adequacy

While high agreement was found to exist between the facility adequacy measures of community groups A and B, almost no agreement was observed between the community, GRD and PRAB. This finding suggests the need for a thorough assessment of the recreational facility needs of the Gainesville community prior to the expansion of recreational facilities.

Input-output relationship

Although the application of regression analysis failed to reveal satisfactory relationships between output quantity and quality and labor inputs, a conclusion that such analysis is inappropriate or that there are no relationships would be premature. "Better" measures of labor input and quantity of output should be obtained and longitudinal as well as crossectional analysis should be performed. Furthermore other input measures (e.g. materials) should be obtained and incorporated into the analyses.

Overall conclusions

The successful development of the performance measurement model for the GRD and the collection and analysis of the data required for its implementation provide evidence that NFP performance measurement is feasible. The model developed appears to meet Knighton's (1972)

requirement for a system which permits the matching of operating expenses with information that provided an indication of public benefits.

Future Applications of Model

Usefulness of model information

The usefulness of the information produced by the model was assumed by the researcher. Now that the information has been obtained, its actual usefulness to Gainesville decision makers needs to be assessed. This assessment should encompass both the cost of and value (in use) of the specific items of model information. The groups presently identified for participation in the evaluation of the model's usefulness are the GRD, PRAB, City Manager's Office and the Gainesville City Commission.

Replication of model

Because of resource constraints (both money and time), numerous limitations were imposed on the data collected. Therefore, this research needs to be replicated with sufficient resources to remove these limitations. With adequate resources, the following improvements are believed to be possible:

1. Program identification would be exhaustive.

2. Having identified all programs, an information system to collect the model data would be implemented. This would eliminate the need to rely on estimates for objective input and output measures. The data would be collected for several years (longitudinal study) and therefore input-output relations could be examined for individual programs as well as across programs.

3. The opinions of a <u>representative</u> sample of community members would be obtained through use of second requests and interviews. A random sample of holdouts would be contacted and <u>paid</u> to be interviewed.

4. The opinions of known program users (based on 2 above) would be obtained for comparison with opinions in 3 above.

5. Program-specific as well as general program quality measures would be obtained.

6. Contextual effects would be controlled for. More than one method (in order to identify methods variance in the multivariable-multirater matrix) would be used to generate measures of program importance and quality.

Delphi

The Delphi questionnaires proved to be a satisfactory means of obtaining opinions from GRD supervisors and PRAB members. Several participants did, however, question the information content of the statistical feedback and in future research experimentation with both statistical and verbal feedback is recommended.

The reasonable distribution criterion for group value judgements can be used to identify problem areas--for those activities about which disagreement exists (bimodal distribution), an investigation can be undertaken to determine if the differences of opinion are due to (1) semantics, (2) values, or (3) the existence of different sets of equally valid facts. Such a management-by-exception tool should be of value to decision makers.

Multivariable-Multirater Matrix

The multivariable-multirater matrix was found to be very useful in assessing the validity of measures of program importance and quality. If only convergent validity had been assessed, the high correlations between importance and quality would not have been noted and erroneous conclusions concerning the validity of the measures might have been made.

Because only one method (5 category rating scale) was used for the measurement of importance and quality, it was not possible to identify the amount of method variance. In future research more than 1 method as well as more than 1 variable and rater group should be used.

Reliance on Self-Evaluations

While self-evaluation studies may provide much valuable information, the results of this research indicate that exclusive reliance on such studies to assess an entity's performance is unjustified and may actually be misleading (as in the case of recreation facility adequacy). To the extent possible, self-evaluations should be augmented by outside appraisals. Furthermore, in order to assure the objectivity of the evaluations, an agency independent of the entity being evaluated (in the case of the GRD, a separate program evaluation department in the City government could be used) should be responsible for validating the information obtained.

The comparison of self with outside evaluations can help decison makers to identify areas in need of more thorough investigation. An

example, from this research, will serve to illustrate such usage.
Although the adult flag football program received the lowest importance
rating awarded by the community, it was rated high in importance by
the GRD. An investigation of the program might reveal that its cost
is completely absorbed by user fees and therefore the program is no
burden to the general community. If, however, it were found that the
program cost is borne primarily by the community and that only a few
members of the community benefit from its provision, then the GRD should
have to demonstrate why the general community (through taxes) should
bear the cost of this program.

APPENDIX A

PROGRAMS AND FACILITIES OF
THE GAINESVILLE RECREATION DEPARTMENT

PROGRAMS PROVIDED BY
AQUATICS DIVISION

Youth swim lessons

Adult swim lessons

Water safety and lifesaving instruction

Springboard diving lessons

Water ballet

Skin & scuba diving lessons

Competitive swimming

Public swimming

Swim meets

PROGRAMS PROVIDED BY
ATHLETICS DIVISION

Tennis lessons

Tennis tournaments

Archery activities

Youth basketball

Adult basketball

Youth football

Adult flag football

Youth baseball

Adult softball

Girl's softball

Racquetball tournaments

Track and field day

Golf lessons

Women's volleyball

Summer track program

PROGRAMS PROVIDED BY
CENTERS DIVISION

Youth ceramics

Adult ceramics

Youth arts & crafts

Adult arts & crafts

Pre-school training

Tumbling lessons

Baton lessons

Square dance lessons

Cooking lessons

Teen nutrition lessons

Modern dance lessons

Adult exercise lessons

Duplicate bridge

Recreation center dances

Recreation center games

Wrestling lessons

Senior citizen activities

Sewing lessons

PROGRAMS PROVIDED BY
PLAYGROUNDS DIVISION

Adult gymnastics

Youth gymnastics

Easter egg hunt

Bowling lessons

Camping skills instruction

Drama workshop & play

Children's art display

Cheerleading clinic

Supervised playground activities

OTHER

Park and picnic facilities

Racquetball facilities

Recreation center facilities

Tennis facilities

FACILITIES MAINTAINED BY
GAINESVILLE RECREATION DEPARTMENT

Archery range

Baseball and softball fields

Park and picnic areas

Playgrounds

Racquetball and handball courts

Recreation centers

Swimming pools

Tennis courts

APPENDIX B

COMMUNITY SURVEY QUESTIONNAIRES
A AND B

PART 1
GENERAL INFORMATION

To be sure that the opinions of the entire community are counted, we need to know some things about you. Please answer the following questions by placing an X in the space next to the answer that best describes you. Please answer all questions.

1. Sex: (1)Male ____ (2)Female ____

2. Age: ____(1)16-20 ____(2)21-30
   ____(3)31-50 ____(4)51-65
   ____(5)over 65

3. Highest level of education:
   ____(1)Grammar school
   ____(2)High school diploma
   ____(3)Some college
   ____(4)College degree

4. Married: (1)Yes ____ (2)No ____

5. Number living in home:
   ____(1)1   ____(2)2
   ____(3)3-5   ____(4)over 5

6. Family income: ____(1) 0- 5,000
   ____(2) 5,001-10,000
   ____(3) 10,001-15,000
   ____(4) 15,001-20,000
   ____(5) 20,001-30,000
   ____(6) over 30,000

7. College students: (1)Yes ____ (2)No ____

8. Live in Gainesville City Limits:
   (1)Yes ____ (2)No ____

PART 2
IMPORTANCE OF PROGRAMS

We want to find out your opinion of the importance of the following programs to the Gainesville community. We are interested in your opinion of the importance of the programs to the entire community, not just to you and your family. To the right of each program are five (5) boxes which represent different amounts of importance: very low, low, average, high and very high. Please examine each program and mark an X in the box which best describes your opinion. Please make sure you mark one (1) box for each program listed.

PROGRAMS	VERY LOW 1	LOW 2	AVERAGE 3	HIGH 4	VERY HIGH 5
9. Preschool training					
10. Youth swim lessons					
11. Skin & scuba diving lessons					
12. Adult ceramics					
13. Adult flag football					
14. Adult basketball					
15. Park & picnic facilities					
16. Sewing lessons					
17. Summer track					
18. Drama workshop & play					
19. Public swimming					
20. Square dance lessons					
21. Tumbling lessons					
22. Springboard diving lessons					

PART 3
QUALITY OF PROGRAMS

We need your opinion of the quality of the following programs from the point of view of you and your family. The first five (5) boxes to the right of each program represent different degrees of quality: very poor, poor, fair, good and very good. Please examine each program carefully. If you have an opinion of its quality, mark an X in the box which best describes your opinion. If you do not know the quality of the program, mark an X in the last box which is designated no opinion. Please make sure you mark one box for each program listed.

PROGRAMS	VERY POOR 1	POOR 2	FAIR 3	GOOD 4	VERY GOOD 5	NO OPINION 6
23. Square dance lessons						
24. Skin & scuba diving lessons						
25. Summer track						
26. Springboard diving lessons						
27. Adult ceramics						
28. Drama workshop & play						
29. Preschool training						
30. Park & picnic facilities						
31. Adult flag football						
32. Public swimming						
33. Adult basketball						
34. Sewing lessons						
35. Tumbling lessons						
36. Youth swim lessons						

PART 4
PARTICIPATION IN PROGRAMS

Indicate the frequency of participation by you and the members of your family in the following programs. Do this by marking an X in the box which best describes you and your family. Please make sure you mark one box for each program listed.

FREQUENCY OF PARTICIPATION

PROGRAMS	NEVER PARTICIPATE 1	SOMETIMES PARTICIPATE 2	FREQUENTLY PARTICIPATE 3
37. Park & picnic facilities			
38. Adult basketball			
39. Springboard diving lessons			
40. Drama workshop & play			
41. Pre-school training			
42. Adult ceramics			
43. Public swimming			
44. Summer track			
45. Adult flag football			
46. Tumbling			
47. Skin & scuba diving			
48. Sewing lessons			
49. Square dance lessons			
50. Youth swim lessons			

PART 5
ADEQUACY OF FACILITIES AND PROGRAMS

If you believe the following facilities are adequate for the present needs of the community, mark an X in the box designated adequate. If you do not believe they are adequate, mark an X in the box designated inadequate. Please make sure you mark one box for each facility listed.

ADEQUACY

FACILITIES	ADEQUATE 1	INADEQUATE 2
51. Baseball		
52. Archery		
53. Softball		
54. Parks		
55. Recreation centers		
56. Racquetball & handball		
57. Playgrounds		
58. Swimming pools		
59. Tennis		

Please list any additional programs or facilities which you feel are presently needed.

1.
2.
3.
4.

GAINESVILLE RECREATION DEPARTMENT

Dear Resident:

WANT A VOICE IN CITY GOVERNMENT? BE COUNTED IN! CAST YOUR BALLOT on the importance and quality of recreational programs provided by the GAINESVILLE RECREATION DEPARTMENT.

We want your opinions about the importance and quality of our programs. By importance, we mean how much a program contributes to making the Gainesville community a more enjoyable place to live. By quality, we mean how good the program is. If you will take a few minutes to complete this questionnaire, you will help us find out how good our programs are. If you don't answer, your opinions can't be counted.

There are five (5) parts to this questionnaire. Please be sure you complete each part by answering all the questions in each part. To assist us in using the results of the questionnaire, please use an to mark your answers.

Sincerely,

Roy Massey, Director
GAINESVILLE RECREATION DEPARTMENT

PART 1
GENERAL INFORMATION

To be sure that the opinions of the entire community are counted, we need to know some things about you. Please answer the following questions by placing an X in the space next to the answer that best describes you. Please answer all questions.

1. Sex: (1)Male _____ (2)Female _____

2. Age: (1)16-20 _____ (2)21-30 _____
   (3)31-50 _____ (4)51-65 _____
   (5)over 65 _____

3. Highest level of education:
   (1)Grammar school _____
   (2)High school diploma _____
   (3)Some college _____
   (4)College degree _____

4. Married: (1)Yes _____ (2)No _____

5. Number living in home:
   (1) 1 _____   (2) 2 _____
   (3) 3-5 _____ (4) over 5 _____

6. Family income: (1)   0- 5,000 _____
   (2) 5,001-10,000 _____
   (3) 10,001-15,000 _____
   (4) 15,001-20,000 _____
   (5) 20,001-30,000 _____
   (6) over 30,000 _____

7. College student: (1)Yes _____ (2)No _____

8. Live in Gainesville City Limits:
   (1)Yes _____ (2)No _____

PART 2
IMPORTANCE OF PROGRAMS

We want to find out your opinion of the importance of the following programs to the Gainesville community. We are interested in your opinion of the importance of the programs to the entire community, not just to you and your family. To the right of each program are five (5) boxes which represent different amounts of importance: very low, low average, high and very high. Please examine each program and mark an X in the box which best describes your opinion. Please make sure you mark one (1) box for each program listed.

IMPORTANCE

PROGRAMS	VERY LOW 1	LOW 2	AVERAGE 3	HIGH 4	VERY HIGH 5
9. Summer track ...					
10. Modern dance lessons					
11. Youth arts & crafts					
12. Archery activities ...					
13. Youth gymnastics ...					
14. Racquetball facilities ...					
15. Golf lessons ...					
16. Track & field day					
17. Adult ceramics					
18. Recreation center dances					
19. Tumbling lessons ...					
20. Youth ceramics					
21. Springboard diving lessons ...					
22. Wrestling lessons ...					

PART 3
QUALITY OF PROGRAMS

We need your opinion of the quality of the following programs (from the point of view of you and your family). The first five (5) boxes to the right of each program represent different degrees of quality: very poor, poor, fair, good and very good. Please examine each program carefully. If you have an opinion of its quality, mark an X in the box which best describes your opinion. If you do not know the quality of the program, mark an X in the last box which is designated no opinion. Please make sure you mark one box for each program listed.

QUALITY

PROGRAMS	VERY POOR 1	POOR 2	FAIR 3	GOOD 4	VERY GOOD 5	NO OPINION 6
23. Springboard diving lessons ...						
24. Golf lessons						
25. Summer track ...						
26. Youth gymnastics ...						
27. Tumbling lessons						
28. Recreation center dances						
29. Racquetball facilities ...						
30. Youth arts & crafts ...						
31. Wrestling lessons						
32. Modern dance lessons ...						
33. Track & field day ...						
34. Youth ceramics						
35. Archery activities ...						
36. Adult ceramics						

PART 4
PARTICIPATION IN PROGRAMS

Indicate the frequency of participation by you and the members of your family in the following programs. Do this by marking an X in the box which best describes you and your family. Please make sure you mark one box for each program listed.

	FREQUENCY OF PARTICIPATION		
PROGRAMS	NEVER PARTICIPATE 1	SOMETIMES PARTICIPATE 2	FREQUENTLY PARTICIPATE 3
37. Recreation center dances			
38. Modern dance lessons			
39. Youth ceramics			
40. Track & field			
41. Youth arts & crafts day			
42. Youth			
43. Springboard diving lessons			
44. Archery			
45. Golf lessons activities			
46. Tumbling lessons			
47. Racquetball facilities			
48. Wrestling lessons			
49. Summer track			
50. Adult ceramics			

PART 5
ADEQUACY OF FACILITIES AND PROGRAMS

If you believe the following facilities are adequate for the present needs of the community, mark an X in the box designated adequate. If you do not believe they are adequate, mark an X in the box designated inadequate. Please make sure you mark one box for each facility listed.

	ADEQUACY	
FACILITIES	ADEQUATE 1	INADEQUATE 2
51. Baseball		
52. Archery		
53. Softball		
54. Parks		
55. Recreation centers		
56. Racquetball & handball		
57. Playgrounds		
58. Swimming pools		
59. Tennis		

Please list any additional programs or facilities which you feel are presently needed.

1.
2.
3.
4.

GAINESVILLE RECREATION DEPARTMENT

Dear Resident:

WANT A VOICE IN CITY GOVERNMENT? BE COUNTED IN! CAST YOUR BALLOT on the importance and quality of recreational programs provided by the GAINESVILLE RECREATION DEPARTMENT.

We want your opinions about the importance and quality of our programs. By importance, we mean how much a program contributes to making the Gainesville community a more enjoyable place to live. By quality, we mean how good the program is. If you will take a few minutes to complete this questionnaire, you will help us find out how good our programs are. If you don't answer, your opinions can't be counted.

There are five (5) parts to this questionnaire. Please be sure you complete each part by answering all the questions in each part. To assist us in using the results of the questionnaire, please use an X to mark your answers.

Sincerely,

Ray McGivery, Director
GAINESVILLE RECREATION
DEPARTMENT

APPENDIX C

COVER LETTER FROM RESEARCHER
TO COMMUNITY

May 19, 1975

Dear Area Resident:

I am a Ph.D. candidate in Accounting at the University of Florida. With the cooperation of the Gainesville Recreation Department, I am conducting a study, for my doctoral thesis, of the importance and quality of recreational programs provided by the Recreation Department.

A scientifically determined group of residents in the Gainesville community has been selected to represent the opinions of the community. As a member of that group, your opinions will play an important role in helping the Gainesville Recreation Department provide those programs which contribute most to the enjoyment of life in the Gainesville area. By taking a few minutes to complete the enclosed questionnaire, you will help the Gainesville Recreation Department serve you and the community better.

Please read carefully the introductory letter by the Director of the Recreation Department and then complete the questionnaire. Please do not sign the questionnaire. Your identity will remain unknown. A stamped self-addressed envelope is enclosed for your reply.

Thank you for your cooperation.

Sincerely,

Marcus Dunn

MD/meb

Enclosure

APPENDIX D

OBJECTIVE DATA COLLECTION FORMS

Where Held: _____

When Held: _____

Fees paid by participant:
  Instructor fee   _____
  Supplies/materials usage fee _____

Participation:

  Number of classes offered per year   _____
  Number of meetings per class   _____
  Average number of participants per meeting   _____
  Number of hours a meeting lasts   _____

Direct Costs to Gainesville Recreation Department

  Number of instructors per meeting   _____

  Type of instructor and rate of pay:

		Per Hour Pay Rate
Part-time staff	_____	_____
Full-time staff:		_____
Supv. II	_____	_____
Supv. I	_____	_____
Aide II	_____	_____
Aide I	_____	_____
Volunteer	_____	_____

  Supplies/materials cost per class   _____

  Equipment cost per class   _____

  Other direct costs per class:

    _____   _____
    _____   _____
    _____   _____
    _____   _____

Participation

    Number of meetings per activity per year   _____
    Average number of hours each meeting lasts   _____
    Average number of participants per meeting   _____

Direct Costs to Gainesville Recreation Department

Personnel Cost:	Number of Hours Per Meeting	Pay Rate Per Hour
Part-time Staff	_____	_____
Full-time staff:		
Supv. II	_____	_____
Supv. I	_____	_____
Aide II	_____	_____
Aide I	_____	_____
Volunteer	_____	_____

Materials/supplies cost per meeting   _____

Utilities cost per meeting   _____

Maintenance cost per meeting   _____

Other cost per meeting _____   _____
                        _____   _____
                        _____   _____

Participation

    Number of recreational swimmers per year   _____

    Average number of hours of pool use per swimmer
       per visit to pool   _____

Direct Costs to Gainesville Recreation Department

Personnel Cost:	Number	Pay Rate Per Hour	Number of Hours Per Person
Cashier	_____	_____	____
Lifeguard	_____	_____	____
Manager	_____	_____	____
Other	_____	_____	____

Materials/supplies cost per year   _____

Utilities cost per year   _____

Maintenance cost per year   _____

Other costs per year   _____   _____

                       _____   _____

                       _____   _____

                       _____   _____

                       _____   _____

Entry fee per participant: _____   _____

Where held: _____

When held: _____

Participation

    Number of participants   _____
    Average number of hours of participation
       per participant   _____

Spectators

    Number of spectators   _____
    Average number of hours of viewing by spectator _____

Direct Costs to Gainesville Recreation Department

    Personnel:

	Hours worked	Number	Pay Rate
Part-time	_____	_____	_____
Full-time			
Supv. II	_____	_____	_____
Supv. I	_____	_____	_____
Aide II	_____	_____	_____
Aide I	_____	_____	_____
Volunteer	_____	_____	_____

    Supplies/materials cost   _____

    Equipment cost   _____

    Trophies, prizes, etc.   _____

    Other (list) _____   _____

              _____   _____

              _____   _____

Entry fee: _____   Time played: _____

Where played _____ % of games requiring lights _____

                 _____

                 _____

                 _____

                 _____

Participation

    Players:
        Number of teams   _____
        Number of games played by each team   _____
        Actual number of players per team   _____
        Average number of hours a game lasts   _____

    Number of teams   _____
    Number of practice sessions per team   _____
    Average number of players per team practice session _____
    Average number of hours a practice session lasts   _____

    Spectators:
        Average number of spectators per game   _____

    Volunteers:
        Coaches--number of hours per team per season   _____
        Other volunteer hours per team per season   _____

Direct Costs to Gainesville Recreation Department

    Personnel cost per game:

	Number	Pay Rate (Designate per hour or game)
Referee/umpire	_____	_____
Scorekeeper	_____	_____
Timekeeper	_____	_____
Field Supervisor	_____	_____
Other	_____	_____

    Equipment cost per team   _____

    Uniform cost per team   _____

    Other direct (non-maintenance) cost per team   _____

Maintenance cost per game   _____

Utilities cost per game   _____

APPENDIX E

DELPHI QUESTIONNAIRE

All 4 parts of the questionnaire (Part 1 - Part 4) were used for
round 1. For round 2 only Part 1 and Part 2 were used. The section
for feedback information--"Quartiles of Responses"--was used on round
2 only.

THE DELPHI TECHNIQUE

"The Delphi technique, developed by the Rand Corporation over 20 years ago, is a method for eliciting and refining group judgements. The rationale for the procedures is primarily the age-old adage 'Two heads are better than one,' when the issue is one where exact knowledge is not available. The procedures have three features: (1) Anonymous response-opinions of members of the group are obtained by formal questionnaire. (2) Iteration and controlled feedback-iteration is effected by a systematic exercise conducted in several iterations, with carefully controlled feedback between rounds. (3) Statistical group response-the group opinion is defined as an appropriate aggregate of individual opinions on the final round. These features are designed to minimize the biasing effects of dominant individuals, of irrelevant communications, and of group pressure toward conformity."

The technique has received extensive testing by the Rand Corporation and others and has proven superior to face-to-face group discussion and to individuals acting alone. The technique has been extensively applied by both industry and government. It has been used primarily to forecast future events, define and rank organizational goals, and measure the value and quality of services provided by not-for-profit organizations.

Dear

     This is the first in a series of questionnaires designed to obtain the opinions of the members of the Gainesville Recreation Advisory Board as to the importance and quality of recreation programs provided by the Gainesville Recreation Department. In each of the questionnaires, you will be asked to express your opinion of the <u>im-portance</u> (Part 1) and <u>quality</u> (Part 2) of recreation programs. In addition, for the first questionnaire <u>only</u> (enclosed), you are asked to

     1. Indicate the frequency of participation by you and the members of your family in the recreation programs (Part 3).

     2. Indicate your opinion of the adequacy of recreation facilities (Part 4).

     The instructions for completing each part of the questionnaire are given at the beginning of that part. If you have any questions concerning how to complete the questionnaire, please contact me.

     The opinions of each board member will be combined to produce a group judgement (group response) on the importance and quality of each program. On the second questionnaire, you will be provided with the group response for each program, the range of the middle 50 percent of the individual responses, and your response on the previous questionnaire. In light of this feedback information you will be asked to reconsider your previous response. (The columns crossed out in the first questionnaire are for the feedback information).

     Since <u>anonymity</u> is a very important part of a Delphi exercise, please do not discuss the questionnaire with other board members or Recreation Department administrators (who are also participating in a separate Delphi exercise) until you complete the final questionnaire. I am interested in your opinions uninfluenced by other board members or recreation administrators.

     Because your opinions play an important and vital role in my study, they are extremely valuable. Thank-you very much for your cooperation.

     If you have any questions, please contact me at 372-7279.

                      Sincerely,

                      Mark Dunn

PART 1
IMPORTANCE OF PROGRAMS

We want to find out your opinion of the <u>importance</u> of the following programs to the Gainesville community. We are interested in your opinion of the importance of the programs to the entire community, not just to you and your family. To the right of each program are five (5) boxes which represent different amounts of importance: <u>very low</u>, <u>low, average, high</u> and <u>very high</u>. Please examine each program and mark an X in the box which best describes your opinion. <u>Please make sure you mark one (1) box for each program listed.</u>

PROGRAMS	IMPORTANCE					QUARTILES OF RESPONSES			
	VERY LOW	LOW	AVERAGE	HIGH	VERY HIGH	PRIOR RESPONSE	LOW	MEDIAN	HIGH
1. Youth baseball									
2. Drama workshop & play...............									
3. Park & picnic facilities........									
4. Teen nutrition lessons...........									
5. Golf lessons									
6. Pre school training..........									
7. Recreation center games..............									
8. Youth football									
9. Springboard diving lessons.....									
10. Tennis facilities.........									
11. Bowling lessons............									
12. Tumbling lessons............									
13. Modern dance lessons............									
14. Youth gymnastics.........									
15. Baton lessons............									

| | | IMPORTANCE | | | | | | QUARTILES OF RESPONSES | | | |
	VERY LOW	LOW	AVERAGE	HIGH	VERY HIGH		PRIOR RESPONSE	LOW	MEDIAN	HIGH
16. Tennis lessons............										
17. Youth basketball.........										
18. Summer track..............										
19. Adult softball...........										
20. Senior citizen activities.........										
21. Racquetball tournaments........										
22. Sewing lessons............										
23. Track & field day................										
24. Girl's softball............										
25. Adult gymnastics.........										
26. Duplicate bridge.............										
27. Youth swim lessons............										
28. Swim meets..............										
29. Water safety & lifesaving inst....										
30. Wrestling lessons............										
31. Adult ceramics...........										
32. Water ballet.............										
33. Archery activities.........										
34. Supervised playground activities.........										
35. Recreation center dances............										
36. Adult flag football...........										

	IMPORTANCE						QUARTILES OF RESPONSES			
	VERY LOW	LOW	AVERAGE	HIGH	VERY HIGH		PRIOR RESPONSE	LOW	MEDIAN	HIGH
37. Youth ceramics.............										
38. Women's volleyball..........										
39. Square dance lessons.............										
40. Adult exercise lessons.............										
41. Adult arts & crafts..............										
42. Competitive swimming............										
43. Racquetball facilities..........										
44. Public swimming............										
45. Camping skills inst...............										
46. Youth arts & crafts..............										
47. Cooking lessons.............										
48. Easter egg hunt................										
49. Cheerleading lessons.............										
50. Skin & scuba diving lessons......										
51. Tennis tournaments.........										
52. Art display..............										
53. Recreation center facilities..........										
54. Adult basketball..........										
55. Adult swim lessons.............										

PART 2
QUALITY OF PROGRAMS

We need your opinion of the quality of the following programs (from the point of view of you and your family). The first five (5) boxes to the right of each program represent different degrees of quality: very poor, poor, fair, good and very good. Please examine each program carefully. If you have an opinion of its quality, mark an X in the box which best describes your opinion. If you do not know the quality of the program, mark an X in the box which is designated no opinion. Please make sure you mark one box for each program listed.

PROGRAMS	VERY POOR	POOR	FAIR	GOOD	VERY GOOD	NO OPINION	PRIOR RESPONSE	LOW	MEDIAN	HIGH
QUALITY							QUARTILES OF RESPONSES			
56. Water safety & lifesaving inst....										
57. Recreation center games..............										
58. Racquetball facilities.........										
59. Youth gymnastics.........										
60. Public swimming..........										
61. Skin & scuba diving lessons...........										
62. Easter egg hunt...............										
63. Recreation center dances.............										
64. Senior citizen activities.........										
65. Tennis tournaments........										
66. Youth ceramics...........										
67. Tumbling lessons...........										
68. Competitive swimming...........										
69. Campign skills inst..............										

	QUALITY							QUARTILES OF RESPONSES			
	VERY POOR	POOR	FAIR	GOOD	VERY GOOD	NO OPINION		PRIOR RESPONSE	LOW	MEDIAN	HIGH
70. Youth basketball........											
71. Adult swim lessons..........											
72. Baton lessons											
73. Youth football..........											
74. Duplicate bridge...........											
75. Youth swim lessons..........											
76. Youth arts & crafts.........											
77. Wrestling lessons..........											
78. Track & field day..............											
79. Adult arts & crafts.........											
80. Springboard diving lessons....											
81. Youth basketball........											
82. Summer track............											
83. Tennis facilities........											
84. Adult softball..........											
85. Pre-school training..........											
86. Adult flag football..........											
87. Bowling lessons											
88. Tennis lessons..........											
89. Adult basketball........											
90. Archery activities........											

		QUALITY						QUARTILES OF RESPONSES			
	VERY POOR	POOR	FAIR	GOOD	VERY GOOD	NO OPINION	PRIOR RESPONSE	LOW	MEDIAN	HIGH
91 Modern dance lessons............											
92. Recreation center facilities.........											
93. Square dance lessons............											
94. Teen nutrition lessons............											
95. Sewing lessons............											
96. Girl's softball...........											
97. Park & picnic facilities.........											
98. Water ballet.............											
99. Drama workshop & play.............											
100. Adult ceramics...........											
101. Women's volleyball											
102. Supervised play-ground activities..											
103. Adult exercise lessons............											
104. Cooking lessons............											
105. Art display............											
106. Golf lessons............											
107. Adult gymnastics.........											
108. Cheerleading lessons............											
109. Racquetball tournaments........											
110. Swim meets..............											

PART 3
PARTICIPATION IN PROGRAMS

Indicate the frequency of participation by you and the members of your family in the following programs. Do this by marking an X in the box which best describes you and your family. Please make sure you mark one box for each program listed.

FREQUENCY OF
PARTICIPATION

PROGRAMS	NEVER PARTICIPATE	SOMETIME PARTICIPATE	FREQUENTLY PARTICIPATE
111. Swim meets			
112. Youth ceramics			
113. Youth baseball			
114. Archery activities			
115. Skin & scuba diving lessons			
116. Youth swim lessons			
117. Easter egg hunt			
118. Tennis facilities			
119. Golf lessons			
120. Water ballet			
121. Teen nutrition lessons			
122. Racquetball tournaments			
123. Sewing lessons			
124. Park & picnic facilities			
125. Competitive swimming			
126. Racquetball facilities			

	FREQUENCY OF PARTICIPATION			
	NEVER PARTICIPATE.	SOMETIMES PARTICIPATE.	PARTICIPATE.	FREQUENTLY PARTICIPATE.
127. Public swimming.................				
128. Square dance lessons...................				
129. Youth arts & crafts...................				
130. Women's volleyball...............				
131. Track & field day.....................				
132. Adult swim lessons...................				
133. Adult gymnastics...............				
134. Tumbling lessons...................				
135. Recreation center dances...................				
136. Tennis tournaments..............				
137. Youth basketball...............				
138. Springboard diving lessons...................				
139. Camping skills inst.....................				
140. Modern dance lessons...................				
141. Baton lessons...................				
142. Cooking lessons...................				
143. Adult flag football.................				
144. Tennis lessons...................				
145. Supervised playground activities...............				
146. Adult ceramics.................				

	FREQUENCY OF PARTICIPATION		
	NEVER PARTICIPATE.	SOMETIMES PARTICIPATE.	FREQUENTLY PARTICIPATE
147. Pre school training................			
148. Youth gymnastics..............			
149. Wrestling lessons................			
150. Water safety & lifesaving inst.........			
151. Summer track...................			
152. Girl's softball................			
153. Adult exercise lessons................			
154. Cheerleading lessons................			
155. Duplicate bridge..................			
156. Bowling lessons................			
157. Adult softball...............			
158. Adult arts & crafts.................			
159. Art display.................			
160. Adult basketball			
161. Youth football................			
162. Recreation center games...................			
163. Senior citizen activities..............			
164. Drama workshop & play...................			
165. Recreation center facilities..............			

PART 4
ADEQUACY OF FACILITIES AND PROGRAMS

If you believe the following facilities are adequate for the present needs of the community, mark an X in the box designated adequate. If you do not believe they are adequate, mark an X in the box designated inadequate. Please make sure you mark one box for each facility listed.

ADEQUACY

FACILITIES	ADEQUATE	INADEQUATE
166. Baseball	
167. Archery	
168. Softball	
169. Parks	
170. Recreation centers.....	
171. Racquetball & handball....	
172. Playgrounds	
173. Swimming pools.......	
174. Tennis	

Please list any additional programs or facilities which you feel are presently needed.

1. _____

2. _____

3. _____

4. _____

APPENDIX F

COVER LETTERS FOR ROUND ONE
DELPHI QUESTIONNIARE

1. FROM DIRECTOR OF GRD TO GRD SUPERVISORS

2. FROM DIRECTOR OF GRD AND CHAIRMAN OF PUBLIC
    RECREATION ADVISORY BOARD TO PRAB MEMBERS

May 28, 1975

Dear

As you know, Mark Dunn, a Ph.D. candidate in accounting at the University of Florida, is conducting a study of recreation programs provided by our Department. One part of his study involves judgements, by informed individuals, of the importance and quality of recreation programs. Your opinions of the importance and quality of our programs will be solicited by a series of questionnaires (which you helped pre-test). These opinions will play an important role in Mr. Dunn's study.

Although the programs presented for your evaluation are believed to be representative of our activities, we recognize that some may have been omitted.

Mr. Dunn will personally process and analyze each questionnaire. In view of the fact that Mr. Dunn has to analyze the information from the first questionnaire and return the results of the analysis to you on the second questionnaire, I would like for you to return the first questionnaire (attached) to my office no later than Friday, June 6, 1975. If you have any questions concerning the questionnaire or the overall study, please contact Mr. Dunn at 372-7279.

Since Mr. Dunn wants your own opinions uninfluenced by other members of the Department, please do not discuss the questionnaire or your responses with any other member of this Department until you have completed the final questionnaire.

I believe that Mr. Dunn's study will provide valuable insight into our activities and how they can be improved. Thank-you for your co-operation.

Sincerely,

Albert R. Massey, Director of Recreation
GAINESVILLE RECREATION DEPARTMENT

May 23, 1975

Dear

As you know, Mark Dunn, a Ph.D. candidate in accounting at the University of Florida, is conducting a study of recreational programs provided by the Gainesville Recreation Department. One part of his study involves judgements, by informed individuals, of the importance and quality of recreation programs. As a member of the Gainesville Recreation Advisory Board, you are considerably more aware of recreation activities than the average citizen. Therefore, your opinions, which will be solicited by a series of questionnaires, will play an important role in Mr. Dunn's study.

Although the programs presented for your evaluation are believed to be representative of the major recreation activities, we recognize that some may have been omitted.

Mr. Dunn will personally process and analyze each questionnaire. In view of the fact that Mr. Dunn has to analyze the information from the first questionnaire and return the results of the analysis to you on the second questionnaire, we would appreciate your returning the first questionnaire (enclosed) to Mr. Dunn by June 6, 1975. A stamped, self-addressed envelope is enclosed for this purpose. If you have any questions concerning either the questionnaire or the overall study, please contact Mr. Dunn at 372-7279.

Enclosed please find

    1. A description of the Delphi technique
    2. Instructions for the first questionnaire
    3. The first questionnaire

We believe that Mr. Dunn's study will provide insight into the Recreation Department's activities and how they can be improved. Therefore, we encourage and thank-you for your cooperation.

Sincerely,

Hal Ingman, Chairman
GAINESVILLE RECREATION ADVISORY BOARD

Albert R. Massey, Director of Recreation
GAINESVILLE RECREATION DEPARTMENT

APPENDIX G

COVER LETTER AND INSTRUCTIONS FOR
ROUND TWO DELPHI QUESTIONNAIRE

Dear

      Thank-you very much for your cooperation in completing the first questionnaire. In this, the final questionnaire, you are asked to reconsider your previous responses after taking into account some feedback information of the responses of other participants on the first questionnaire. Taking this information into account, you may revise your opinion where you feel it is appropriate. <u>Please rate all programs whether you change your opinion or not</u>.

      The feedback information is listed in the columns designated "Quartiles of Responses."* The numbers in these columns correspond to the numbers at the top of the columns used for rating the Quality and Importance of programs. Thus for "Importance," we have

	Numerical Rating
Very Low	1
Low	2
Average	3
High	4
Very High	5

and for "Quality," we have

Very Poor	1
Poor	2
Fair	3
Good	4
Very Good	5
No Opinion	6

      The type of feedback information is as follows:

      1. Your <u>own response</u> on the first questionnaire is provided in the column labeled "Prior Response." It is written in red ink.

      2. <u>Measures of the group response</u> which are designated "Low," "Median" and "High."

      The "Median" represents the middle (central) response of the group—50% of the responses are less than or equal to the median and 50% of the responses are greater than or equal to the median.

      The number in the "Low" column means that 25% of the responses are less than or equal to this number; the number in the "High" column means that 25% of the responses are greater than or equal to this number. Thus <u>at least 50%</u> of the group responses lie in the interval designated by the "Low" and "High" numbers. Responses outside of the interval represent, in a statistical sense, extreme scores. Some feedback examples follow

*NOTE: "No opinion" responses have been excluded from the feedback information.

Example	Low	Median	High
a	2	3	4
b	4	5	5

In example "a", 50% of the responses are less than or equal to 3 and 50% are greater than or equal to 3. At least 50% of the group responses are contained in the interval 2 to 4. In this example 1 and 5 would represent extreme scores. In example 'b', 50% of the responses are less than or equal to 5 and 50% are greater than or equal to 5. At least 50% of the group responses are contained in the interval 4 to 5. In this example 1, 2 and 3 represent extreme scores.

If you have any questions concerning the feedback information or the questionnaire, please contact me at 372-7279.

The opinions of each member of the group will be combined to produce the final group judgement on the importance and quality of each program. Again, since anonymity is an important part of a Delphi exercise, please do not discuss the questionnaire with other Advisory Board members or Recreation Department Administrators.

Please return the questionnaire as soon as possible.

Sincerely,

Mark Dunn

APPENDIX H

RESIDUAL PLOTS AND
REGRESSION STATISTICS

Figure 13. Plot of standardized residuals against standardized predicted dependent variable for 45 programs.

215
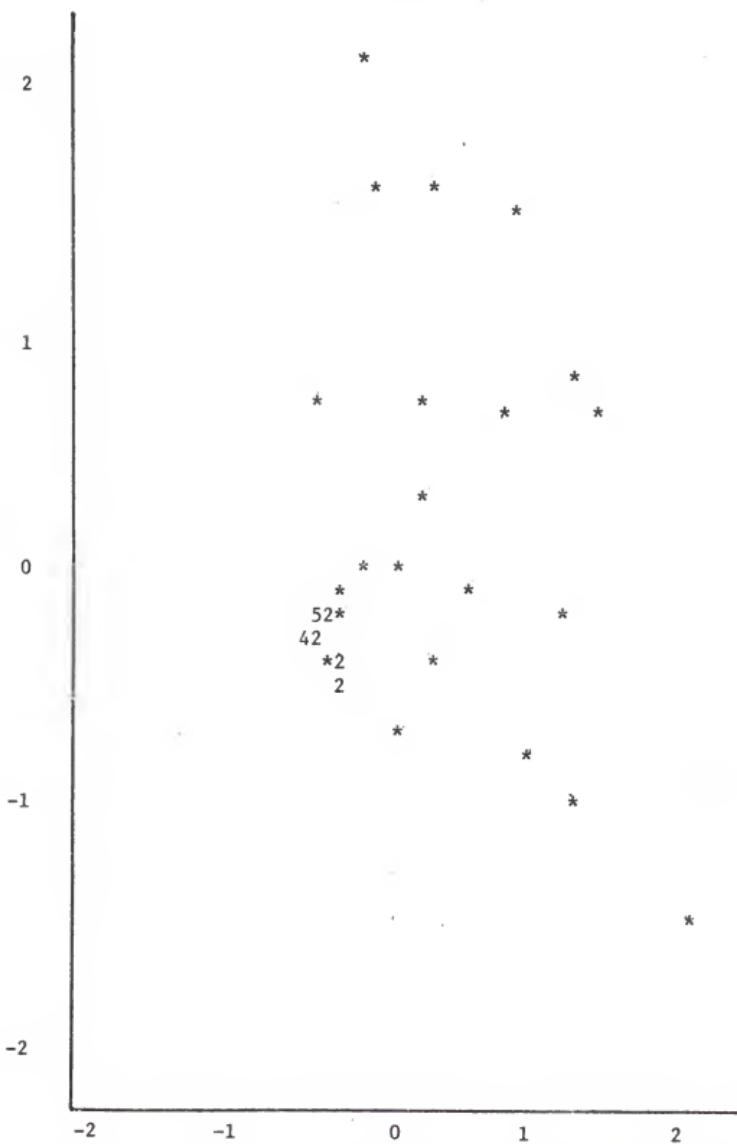
Figure 14. Plot of standardized residuals against standardized
predicted dependent variable for 39 programs.
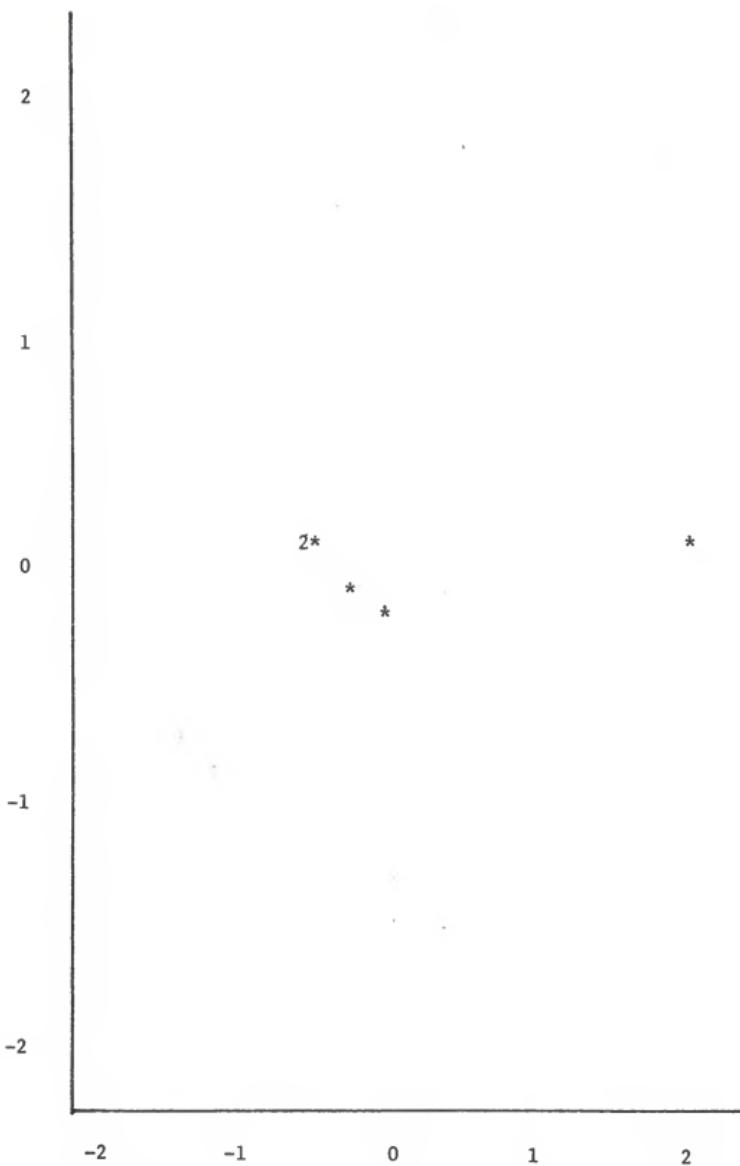
Figure 15. Low quality--plot of standardized residuals against standardized predicted dependent variable for 6 programs.
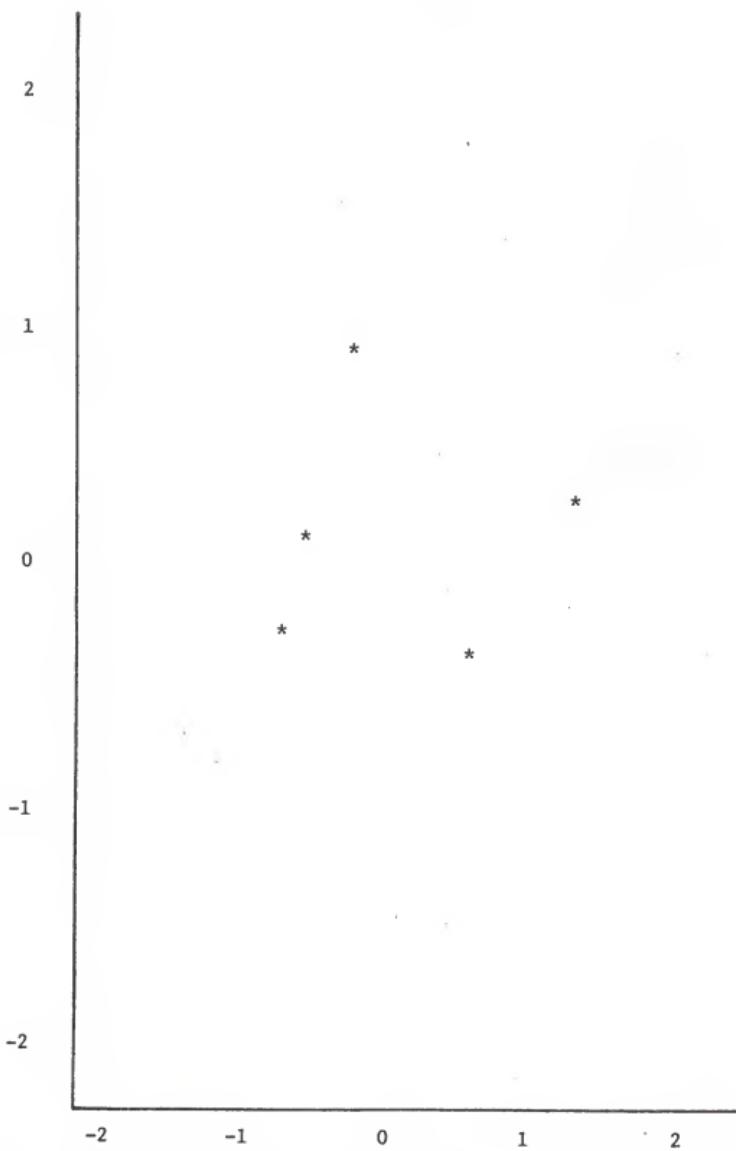
Figure 16. Low quality group--plot of standardized residuals
against stantardized predicted dependent variable
for 5 programs.
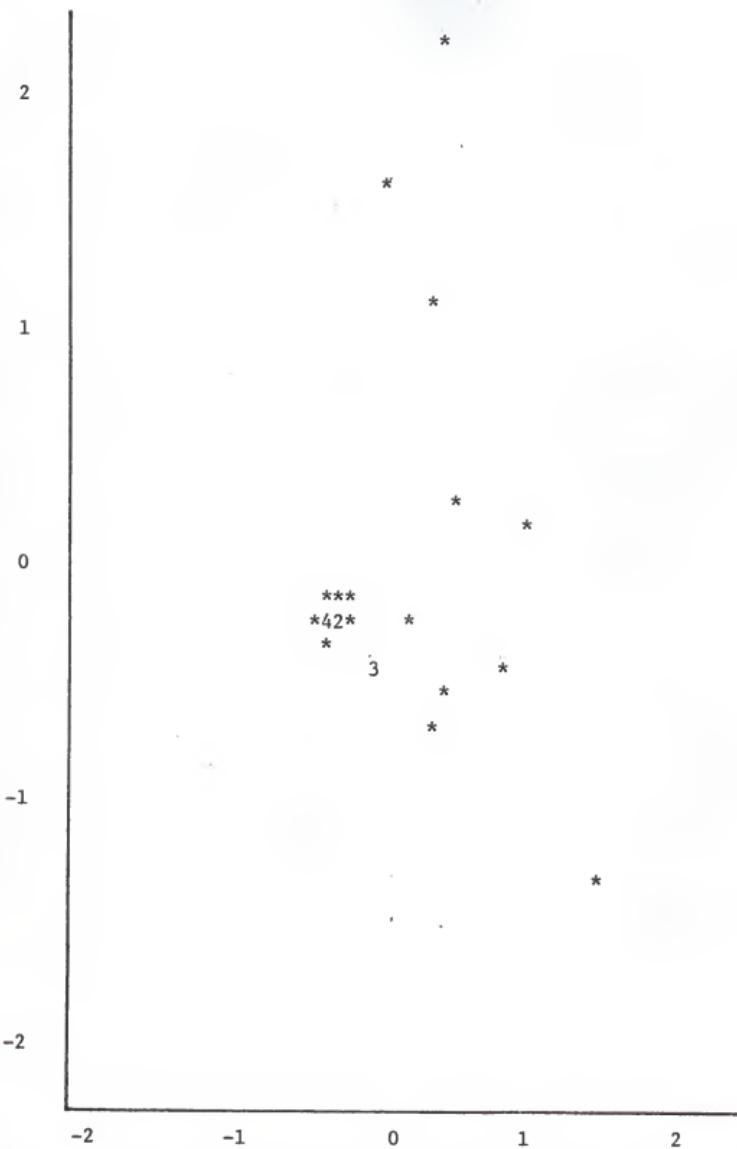
Figure 17. Average quality group--plot of standardized residuals against standardized predicted dependent variable for 25 programs.
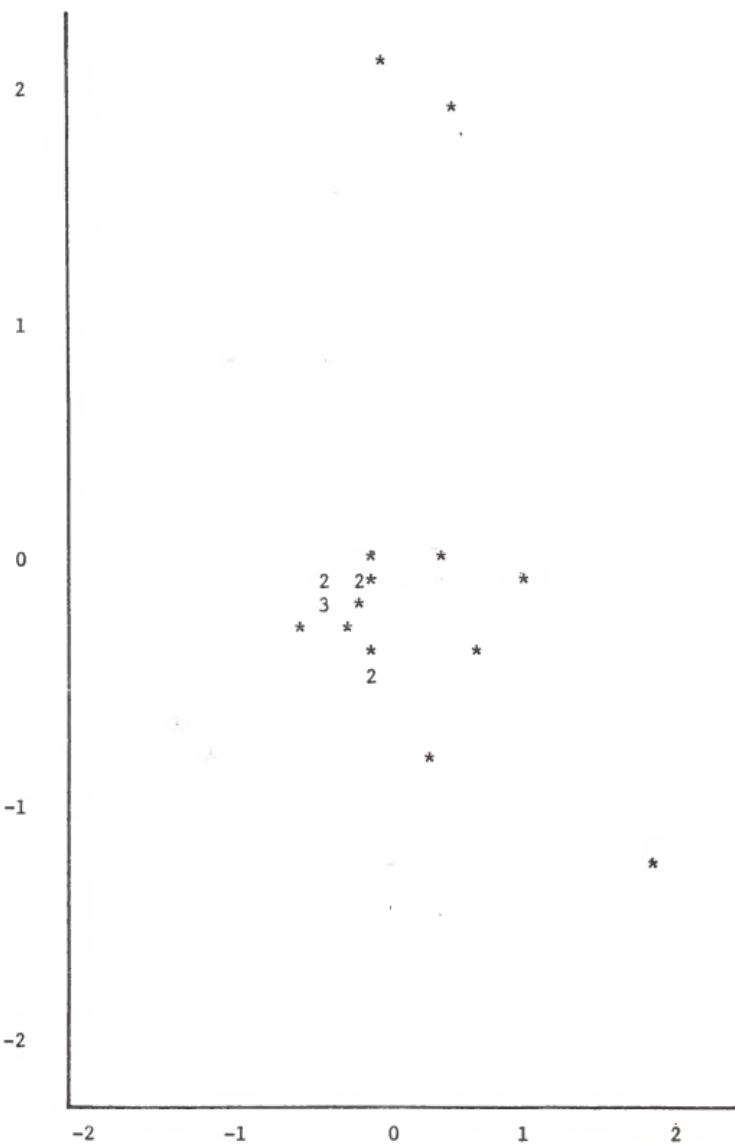
Figure 18. Average quality group--plot of standardized residuals against standardized predicted dependent variable for 24 programs.
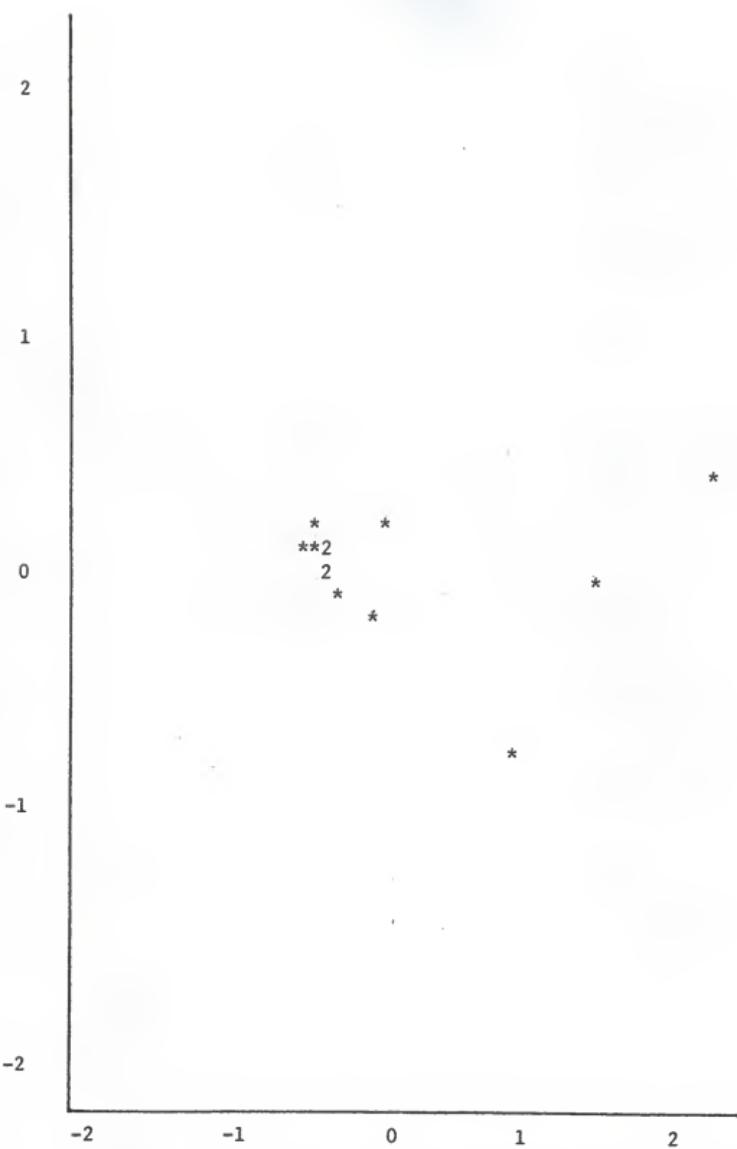
Figure 19. High quality group--plot of standardized residuals against standardized predicted dependent variable for 13 programs.
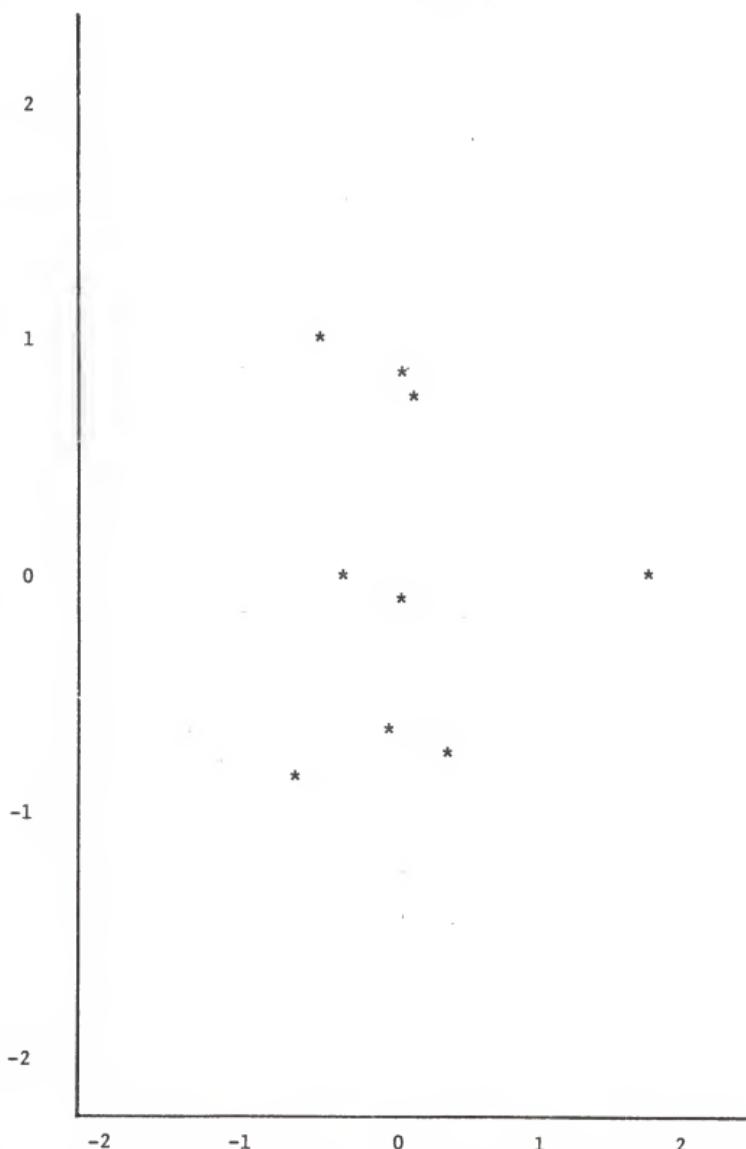
Figure 20. High quality group--plot of standardized residuals against standardized predicted dependent variable for 9 programs.
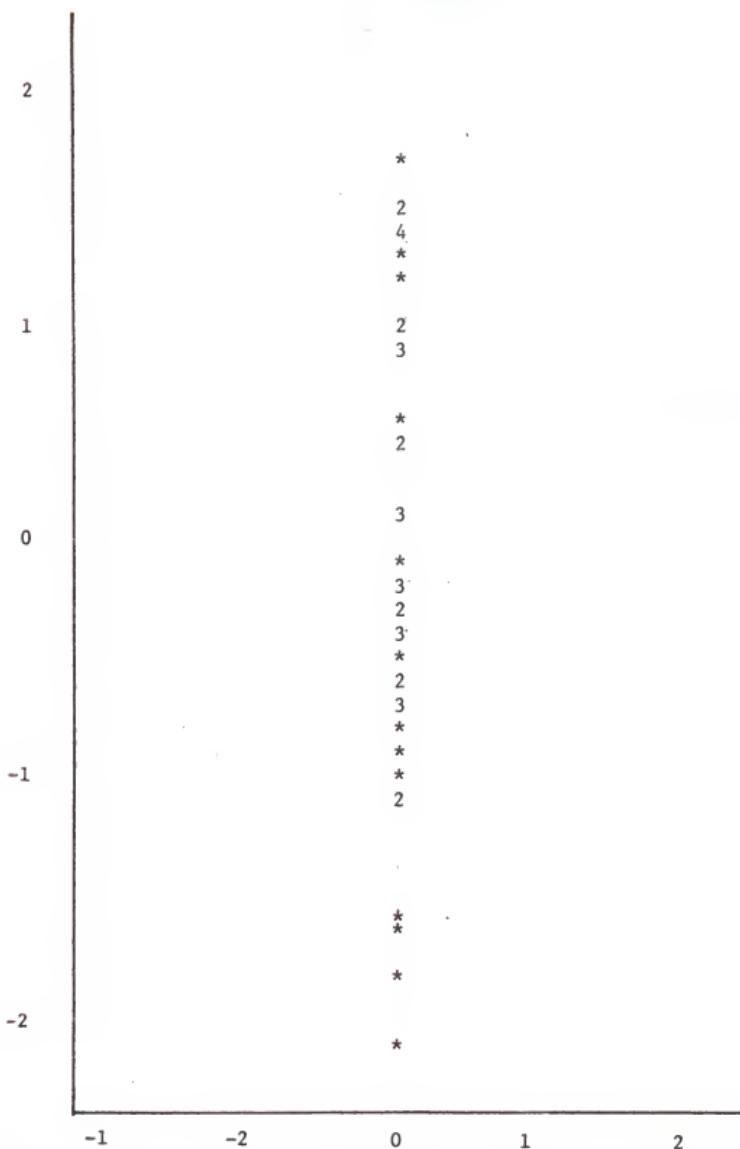
Figure 21. Plot of standardized residuals for quality against the predicted standardized values for quality.

TABLE 31
REGRESSION STATISTICS

Description	n	r	r^2	Standard Error of Estimate	Least Squares Estimates a	Least Squares Estimates b	Standard Error of b	F Value	Level of Significance
E1. U = a + bL									
Overall relationship between participant hours and labor hours	45	.96	.92	9,935	-611	10.75	.47	520	.001
	39	.67	.46	2,756	1,120	7.04	1.27	31	.001
E2. U = a + bL									
Relationship between participant hours and labor hours by level of quality									
Low quality:	6	.99	.98	825	-693	16.15	1.28	159	.001
	5	.86	.74	164	155	3.76	1.27	8.7	NS*
Average quality:	25	.48	.23	4,165	965	9.50	3.62	6.9	.05
	24	.59	.35	2,570	798	7.66	2.25	11.6	.01
High quality:	13	.96	.92	17,603	-5,321	11.18	.98	129	.001
	9	.69	.48	3,247	3,839	3.64	1.43	6.5	.05

TABLE 31--Continued

Description	n	r	r^2	Standard Error of Estimate	Least Squares Estimates a	Least Squares Estimates b	Standard Error of b	F Value	Level of Significance
E3. $Q = a + b \left(\frac{L}{U}\right)$									
Relationship between quality and labor hour--participant hour ratio	45	.01	.000	.52	3.96	-.05	.94	.002	NS *

*Not significant

BIBLIOGRAPHY

American Almanac. New York: Gosset and Dunlap, Inc., 1974.

Bauer, Raymond A. (ed.). Social Indicators, Cambridge, Massachusetts: The M.I.T. Press, 1967.

Bedford, Norton M. Income Determination Theory: An Accounting Framework, Reading, Mass.: Addison-Wesley, 1962.

Beyer, Robert. "The Modern Management Approach to a Program of Social Measurement," Journal of Accountancy, March, 1969, pp. 37-46.

Blalock, Hubert M., Jr. Social Statistics, New York: McGraw-Hill, 1972.

Botner, Stanley B. "Four Years of PPBS: An Appraisal," Public Administration Review, July-August, 1970, pp. 423-431.

Brummet, R. Lee. "Total Performance Measurement," Management Accounting, November, 1973, pp. 11-15.

Campbell, D.T. "Reforms as Experiments," American Psychologist, April, 1969, pp. 409-429.

Campbell, D.T. "Assessing the Impact of Planned Social Change," Paper delivered at the Dartmouth/OECD Seminar, September 13-15, 1974.

Campbell, D.T. and Stanley, J.C. Experimental and Quasi-Experimental Designs for Research, Skokie, Illinois: Rand McNally, 1963.

Campbell, D.T. and Fiske, Donald W. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, March, 1959, pp. 31-105.

Campbell, Robert M. A Methodological Study of the Utilization of Experts in Business Forecasting, Ph.D. Thesis, University of California, Los Angeles, 1966.

Census of Population and Housing--Gainesville, Florida, Standard Metropolitan Statistical Area, 1970. Washington, D.C.: U.S. Bureau of the Census, Department of Commerce, 1972.

Churchill, Neil C. and Stedry, Andrew C. "Extending the Dimensions of Accounting Measurement," Management Services, March-April, 1967, pp. 15-22.

Churchman, C.W. The Systems Approach, New York: Dell Publishing Co., 1968.

Churchman, C. West. "On the Facility, Felicity and Morality of Measuring Social Change," The Accounting Review, January, 1971, pp. 30-35.

Cohn, Elchanan. Public Expenditure Analysis, Lexington, Mass.: D.C. Heath and Co., 1972.

Committee on Accounting Practices of Not-For-Profit Organizations, "Report of the Committee on Accounting Practices of Not-For-Profit Organizations," Supplement to Volume XLVI, The Accounting Review, 1971, pp. 81-163.

Committee on Concepts of Accounting Applicable to the Public Sector. "Report of the Committee on Concepts of Accounting Applicable to the Public Sector," Supplement to Volume XLVII, The Accounting Review, 1972, pp. 77-108.

Committee on Measures of Effectiveness for Social Programs. "Report of the Committee on Measures of Effectiveness for Social Programs," Supplement to Volume XLVII, The Accounting Review, 1972, pp. 337-398.

Committee on Accounting for Human Resources, "Report of the Committee on Accounting for Human Resources," Supplement to Volume XLIX, The Accounting Review, 1974, pp. 114-124.

Committee on Not-For-Profit Organizations. "Report of the Committee on Not-For-Profit Organizations," Supplement to Volume XLIX, The Accounting Review, 1974, pp. 114-124.

Cyert, Richard M. and March, James C. A Behavioral Theory of the Firm, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963.

Dalkey, Norman C. The Delphi Method: An Experimantal Study of Group Opinion, Santa Monica, California: The Rand Corporation, RM-5888-PR, June, 1969

Dalkey, Norman C., Rourke, Daniel L., Lewis, Ralph and Snyder, David. Studies in the Quality of Life-Delphi and Decision Making, Lexington, Massachusetts: D.C. Heath and Company, 1972.

Dorfman, Robert. (ed.). Measuring Benefits of Government Investments, Washington, D.C.: The Brookings Institution, 1965.

Estes, Ralph W. "Socio Economic Accounting and External Diseconomics," The Accounting Review, April, 1972, pp. 284-290.

General Accounting Office. Standards of Audit of Governmental Organizations, Programs, Activities and Functions, Washington, D.C.: U.S. Government Printing Office, 1972.

Goldman, Thomas A., ed. Cost-Effectiveness Analysis, New York: Frederick A. Praeger, 1967.

Gross, Bertram M. "The State of the Nation: Social Systems Accounting," in Social Indicators, edited by Raymond A. Bauer, Cambridge, Mass.: The M.I.T. Press, 1967, pp. 154-271.

Hatry, Harry P. and Dunn, Diana R. Measuring the Effectiveness of Local Government Services-Recreation, Washington, D.C.: The Urban Institute, 1971.

Held, Virginia. "PPBS Comes to Washington," in Planning-Programming-Budgeting: A Systems Approach to Management, edited by Fremont J. Lyden and Ernest G. Miller, Chicago: Markham Publishing Co., 1970, pp. 11-25.

Helmer, Olaf and Gordon, T.J. Report on a Long-Range Forecasting Study, Santa Monica, California: The Rand Corporation, P-2982, 1964.

Helmer, Olaf. Social Technology, New York: Basic Books, Inc. 1966.

Henderson, James M. and Quandt, Richard E. Microeconomic Theory, New York: McGraw-Hill Book Co., 1971.

Henke, Emerson O. "Performance Evaluation for Not-For-Profit Organizations," Journal of Accountancy, June, 1972, pp. 51-55.

Kerlinger, Fred N. Foundations of Behavorial Research, New York: Holt Rinehart and Winston, Inc., 1973.

Knighton, Lennis B. "Performance, Evaluation, and the Matching Concept in Governmental Accounting," Federal Accountant, September, 1969, pp. 95-103.

Knighton, Lennis N. "Accounting for the Benefits of Public Programs," Federal Accountant, March, 1972, pp. 4-19.

Kraus, Richard. Recreation and Leisure in Modern Society, New York: Appleton-Century-Crafts, 1971.

Kraus, Richard and Curtis, Joseph E. Creative Administration in Recreation and Parks, Saint Louis: C.V. Mosby Co., 1973.

Krutilla, J.V. "Welfare Aspects of Benefit-Cost Analysis," in Public Finance--Selected Readings, edited by Helen A. Cameron and William Henderson, New York: Random House, 1966, pp. 175-192.

Lawler, Edward E. "The Multitrait-Multirater Approach to Measuring
    Managerial Job Performance," _Journal of Applied Psychology_,
    October, 1967, pp. 369-381.

Lazarsfeld, Paul F. "Accounting and Social Bookkeeping," reprinted
    in _Accounting in Perspective: Contributions to Accounting
    Thought by Other Disciplines_, edited by R.R. Sterling and
    W.F. Benta, Cincinnati, Ohio: Southwestern Publishing Co., 1971.

Lee, Robert D., and Johnson, Ronald W. _Public Budgeting Systems_,
    Baltimore, Maryland: University Park Press, 1973.

Lindblom, Charles E. "The Science of "Muddling Through,'" _Public
    Administration Review_, Spring, 1959, pp. 79-88.

Lindblom, Charles E. _The Intelligence of Democracy_, New York: The
    Free Press, 1965.

Linowes, David F. "The Accounting Profession and Social Progress,"
    Journal of Accountancy, July, 1973, pp. 32-40.

Lyden, Fremont J. and Miller, Ernest G. (eds.). _Planning-Programming-
    Budgeting: A Systems Approach to Management_, Chicago: Markham
    Publishing Co., 1970.

Mack, Ruth P. and Myers, Sumner. "Outdoor Recreation," in _Measuring
    Benefits of Government Investments_, edited by Robert Dorfman,
    Washington, D.C.: The Brookings Institution, 1965, pp.
    71-116.

Mantel, Samuel J., Jr., Service, Allan L., Reisman, Arnold, Koleski,
    Raymond A., Blum, Arthur, Reich, Robert, Dean, Burton V.,
    Jaffee, Miriam, Reiger, Howard, Ronis, Richard and Rubenstein,
    Judah. _A Social Service Measurement Model_, Technical Memorandum
    No. 283, Department of Operations Research, Case Western
    Reserve University, 1972.

Mobley, Sybil C. "The Challenges of Socio-Economic Accounting," _The
    Accounting Review_, October, 1970, pp. 762-768.

Nie, Norman H., Hull, C. Hadlai, Jenkins, Jean G., Steinbrenner, Karin
    and Bent, Dale H. _Statistical Package for the Social Sciences_,
    second edition, New York: McGraw-Hill Book Co., 1975.

Office of Management and Budget, "Office of Management and Budget
    Transmittal Memorandum No. 38," _Circular A-11_, June 21, 1971.

Parducci, Allen. "Contextual Effects: A Range Frequency Analysis,"
    in _Handbook of Perception, Volume II--Psychophysical Judgement and
    Measurement_, edited by Edward C. Carterette and Morton P. Freidman,
    New York: Academic Press, 1974, pp. 127-141.

Parten, Mildred. Surveys, Polls, and Samples: Practical Procedures, New York: Harper and Brothers, 1950.

Pill, Juri. "The Delphi Method: Substance, Context, A Critique and an Annotated Bibliography," Socio-Economic Planning Science, Volume 5, 1971, pp. 57-71.

Prest, A.R. and Turvey, R. "Cost-Benefit Analysis: A Survey," The Economic Journal, December, 1965, pp. 683-735.

Reisman, Arnold, Mantel, Samuel J., Jr., Dean, Burton V., Eisenberg, Norman C., Markus, Elliot and Service, Allan V. Measurement of Output in a System of Social Agencies, Technical Memorandum No. 188, Department of Operations Research, Case Western Reserve University, 1970.

Rossi, Peter H. "Testing for Success and Failure in Social Action," in Evaluating Social Programs: Theory, Practice and Politics, edited by Peter H. Rossi and Walter Williams, New York: Seminar Press, 1972, pp. 11-49.

Rossi, Peter H. and Williams, Walter (eds.). Evaluating Social Programs: Theory, Practice and Politics, New York: Seminar Press, 1972.

Runkel, Phillip J. and McGrath, Joseph. Research on Human Behavior, New York: Holt, Rinehart and Winston, Inc. 1972.

Schick, Allen. "Systems Politics and Systems Budgeting," Public Administration Review, March-April, 1969, pp. 137-151.

Schick, Allen. "A Death in the Bureaucracy: The Demise of Federal PPB," Public Administration Review, March-April, 1973, pp. 146-156.

Smissen, Van der B. Evaluation and Self-Study of Public Recreation and Park Agencies: A Guide with Standards and Evaluative Criteria, Arlington, Va.: National Recreation and Park Association, 1972.

Terleckyj, Nestor E. "Measuring Progress Towards Social Goals: Some Possibilities at National and Local Levels," Management Science, Volume 16, 1970, pp. B-765-783.

Thompson, Loran T. A Pilot Application of Delphi Technique to the Drug Field: Some Experimental Findings, New York, New York: The New York City Rand Institute, R-1124, 1973.

Tiller, Carl W. "The Demise of PPBS," *Federal Accountant*, June, 1972, pp. 4-13.

Toward A Social Report. U.S. Department of HEW, Washington, D.C.; U.S. Government Printing Office, 1969.

Wennegren, E. Boyd and Fullerton, Herbert H. "Estimating Quality and Location Values of Recreation Resources," *Journal of Leisure Research*, Summer, 1972, pp. 170-183.

Wildavsky, Aaron. *The Politics of the Budgeting Process*, Boston: Little, Brown and Co., 1964.

Wholey, Joseph S., Scanlon, John W., Duffy, H.G., Fukumoto, James S., and Vogt, Leona M. *Federal Evaluation Policy-Analyzing the Effects of Public Programs*, Washington, D.C.: The Urban Institute, 1970.

BIOGRAPHICAL SKETCH

W. Marcus Dunn was born August 17, 1946 in Dallas, Texas. He attended school at Bowie, Texas and graduated from Bowie High School in 1964. Mr. Dunn received a Bachelor of Business Administration degree with a major in accounting from the University of Texas at Arlington, Arlington, Texas in 1968 and a Master of Business Administration degree from the same university in 1970. Mr. Dunn was employed by Arthur Young and Company, Fort Worth, Texas, as a staff accountant from January, 1970 to August, 1972. In September, 1972 he entered the Ph.D. program in business administration at the University of Florida. While working on his Ph.D., Mr. Dunn was associated with the CPA firm of Moore, Cobb and Company. In 1974 Mr. Dunn received an honorary discharge as a staff sergeant from the U.S.A.R. He is presently an Assistant Professor of Accounting at the University of Houston, main campus.

Mr. Dunn is married to the former Diane Mary Murray of Hurst, Texas. They have one daughter, Kristen, who is four years of age.

Mr. Dunn received his CPA certificate (Texas) in 1971 and is a member of the American Institute of Certified Public Accountants, Texas Society of Certified Public Accountants and the American Accounting Association. He is also a member of the honorary societies of Beta Gamma Sigma and Phi Kappa Phi.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
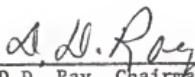
D.D. Ray, Chairman
Professor of Accounting

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
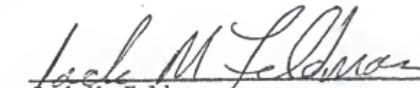
Jack M. Feldman
Associate Professor of Management

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

R. Blaine Roberts
Associate Professor of Economics

This dissertation was submitted to the Graduate Faculty of the Department of Accounting in the College of Business Administration and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1976

Dean, Graduate School